

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2004, 21:23:37 ; Search time 94 seconds  
(without alignments)  
24.047 Million cell updates/sec

Title: US-09-660-302E-1  
Perfect score: 25  
Sequence: 1 XEFIXDX 8

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	8	2	AA32790 Binding s
2	25	100.0	12	2	AA32800 Immunoglob
3	25	100.0	12	2	AA32793 Growth ho
4	25	100.0	12	2	AA32789 Growth ho
5	25	100.0	12	2	AA32794 Chicken g
6	25	100.0	13	6	ABU03293 Human exp
7	25	100.0	16	2	AAW33004 Murine DP
8	25	100.0	36	5	ABP42844 Human ova
9	25	100.0	37	2	AA311253 Streptoco
10	25	100.0	45	2	AA336319 Human nov
11	25	100.0	45	6	ADAL1680 Human nov
12	25	100.0	48	6	ABP72498 Phospholi
13	25	100.0	54	5	ABG80735 C-type le
14	25	100.0	56	4	AA316692 Peptide #
15	25	100.0	56	4	AB35677 Peptide #
16	25	100.0	56	4	AA329179 Peptide #
17	25	100.0	56	4	AB330514 Peptide #
18	25	100.0	56	4	AB21106 Protein #
19	25	100.0	56	4	AA368870 Human bon
20	25	100.0	56	4	AA356491 Human liv
21	25	100.0	56	4	ABG50531 Human pep
22	25	100.0	56	4	AA304409 Peptide #
23	25	100.0	56	5	ABG38451 Human sec
24	25	100.0	57	3	AA302282 Human nov
25	25	100.0	57	3	AA310476 Arabidops

26	25	100.0	57	3	AA324529 Arabidops
27	25	100.0	58	2	AAW90992 Oropouche
28	25	100.0	60	4	AB43221 Peptide #
29	25	100.0	60	4	AA37061 Peptide #
30	25	100.0	60	4	AAW76953 Human bon
31	25	100.0	60	4	AAW64128 Human liv
32	25	100.0	60	4	ABG58615 Human pep
33	25	100.0	61	3	AA334535 Arabidops
34	25	100.0	61	3	AA334535 Arabidops
35	25	100.0	64	4	ABG00521 Novel hum
36	25	100.0	66	3	AA324528 Arabidops
37	25	100.0	66	3	AA310475 Arabidops
38	25	100.0	67	6	ABP80025 N. gonorr
39	25	100.0	68	4	AAU16600 Human nov
40	25	100.0	68	5	ABP64248 Human ORF
41	25	100.0	68	6	ABU55669 Human nov
42	25	100.0	69	3	AA301116 Human sec
43	25	100.0	69	6	ABP77688 N. gonorr
44	25	100.0	72	5	ABU51892 Helicobac
45	25	100.0	73	5	ABP09427 Human ORF
46	25	100.0	74	2	AAW61649 Non-ocula
47	25	100.0	80	5	ABP38159 Staphyloc
48	25	100.0	80	6	ADA36717 Acinetoba
49	25	100.0	82	2	AAW62670 Streptoco
50	25	100.0	82	5	ABP08611 Human ORF
51	25	100.0	83	4	AA83324 Rice isol
52	25	100.0	83	6	ABU08739 Rice isol
53	25	100.0	85	3	AA318654 Arabidops
54	25	100.0	89	5	ABP90813 Herbicida
55	25	100.0	90	1	ABP81887 Sequence
56	25	100.0	94	5	ABP07728 Human ORF
57	25	100.0	95	2	AAW35804 Human end
58	25	100.0	95	2	AA302590 A human m
59	25	100.0	95	3	AA303769 Human end
60	25	100.0	95	3	AA365394 Human 5'
61	25	100.0	95	3	AA322226 Mammoglob
62	25	100.0	95	4	AA31682 An endome
63	25	100.0	95	5	AA302555 Protein o
64	25	100.0	95	5	ABG96366 Human ova
65	25	100.0	95	5	ABG09635 Human end
66	25	100.0	95	6	ABG73169 Human end
67	25	100.0	95	6	ABR47522 Breast ca
68	25	100.0	95	7	ADC78795 Human PRO
69	25	100.0	97	5	ABP01586 Human ORF
70	25	100.0	98	4	ABG63376 Drosophil
71	25	100.0	98	4	AAE03588 Wheat asp
72	25	100.0	99	3	AA334364 Arabidops
73	25	100.0	99	6	ADA36447 Acinetoba
74	25	100.0	101	2	AAW72393 Pathogen
75	25	100.0	103	5	ABP32702 Human ORF
76	25	100.0	105	6	ADA54452 Human PRO
77	25	100.0	106	4	AB317370 Human ner
78	25	100.0	106	6	ABU58228 Wheat str
79	25	100.0	106	7	ADC97504 E. faeciu
80	25	100.0	108	2	AA360038 Human end
81	25	100.0	109	4	AAU02024 B. thurin
82	25	100.0	110	3	AA310474 Arabidops
83	25	100.0	110	3	AA324527 Arabidops
84	25	100.0	110	6	ABR57093 Arabidops
85	25	100.0	113	4	AAU02025 B. thurin
86	25	100.0	114	5	ABG34679 Arabidops
87	25	100.0	114	5	ABG54747 Lactococ
88	25	100.0	116	4	AAU02032 B. thurin
89	25	100.0	116	6	ABU15149 Protein e
90	25	100.0	120	6	ABU1397 Protein e
91	25	100.0	121	3	AA327381 Arabidops
92	25	100.0	123	5	AB31063 Canine lo
93	25	100.0	124	3	AA301456 Human sec
94	25	100.0	124	3	AA34531 Arabidops
95	25	100.0	125	4	AAU16559 Human nov
96	25	100.0	125	6	ABU55628 Human nov
97	25	100.0	125	6	ABU55628 Human nov
98	25	100.0	126	3	AA306674 Arabidops

99 25 100.0 3 AAG35079 Arabidops 172 100.0 165 5 AAM48284 Pheromone  
100 25 100.0 4 AEG03470 Novel hum 173 100.0 166 4 ABG17237 Novel hum  
101 25 100.0 4 AEG03470 Novel hum 174 100.0 167 3 AAG18653 Arabidops  
102 25 100.0 4 AAG16530 Novel hum 175 100.0 167 4 AAU16185 Human nov  
103 25 100.0 3 AAG34872 Arabidops 176 100.0 167 6 ABU55254 Human nov  
104 25 100.0 3 AAG07710 Arabidops 177 100.0 168 3 AAG18652 Arabidops  
105 25 100.0 5 ABB81058 Canine lo 178 100.0 170 3 AAG58217 Arabidops  
106 25 100.0 4 AAG81564 S. epider 179 100.0 171 3 AAG10427 Arabidops  
107 25 100.0 4 AAG42795 Protein e 180 100.0 172 2 AAR12796 Human imm  
108 25 100.0 3 AAY32792 Growth ho 181 100.0 172 2 AAR33377 Soluble C  
109 25 100.0 5 ABBP38346 Staphyloc 182 100.0 172 2 ABU04233 Human exp  
110 25 100.0 3 AAG54766 Arabidops 183 100.0 172 6 ABU04233 Human exp  
111 25 100.0 3 ABB48396 Listeria 184 100.0 172 6 ABU04233 Human exp  
112 25 100.0 3 ABBP98021 Amino aci 185 100.0 175 3 AAG27379 Arabidops  
113 25 100.0 5 ABB81059 Canine lo 186 100.0 176 4 AAG33496 Arabidops  
114 25 100.0 4 AAG79296 Amino aci 187 100.0 177 3 AAG33496 Arabidops  
115 25 100.0 6 ABB81059 Canine lo 188 100.0 178 4 AAU16138 Human nov  
116 25 100.0 4 ABBP38346 Staphyloc 189 100.0 178 6 ABU55207 Mycoplasma  
117 25 100.0 4 ABBP38346 Staphyloc 190 100.0 179 6 ABM69668 Phototrab  
118 25 100.0 4 ABBP38346 Staphyloc 191 100.0 179 6 ABM69668 Phototrab  
119 25 100.0 4 ABBP38346 Staphyloc 192 100.0 179 6 ABM69668 Phototrab  
120 25 100.0 4 ABBP38346 Staphyloc 193 100.0 179 6 ABM69668 Phototrab  
121 25 100.0 4 ABBP38346 Staphyloc 194 100.0 183 5 ABP40491 Staphyloc  
122 25 100.0 4 ABBP38346 Staphyloc 195 100.0 183 5 ABP40491 Staphyloc  
123 25 100.0 4 ABBP38346 Staphyloc 196 100.0 187 5 ABB81061 Canine lo  
124 25 100.0 4 ABBP38346 Staphyloc 197 100.0 187 5 ABB81061 Canine lo  
125 25 100.0 4 ABBP38346 Staphyloc 198 100.0 187 5 ABB81061 Canine lo  
126 25 100.0 4 ABBP38346 Staphyloc 199 100.0 187 5 ABB81061 Canine lo  
127 25 100.0 4 ABBP38346 Staphyloc 200 100.0 187 5 ABB81061 Canine lo  
128 25 100.0 4 ABBP38346 Staphyloc 201 100.0 187 5 ABB81061 Canine lo  
129 25 100.0 4 ABBP38346 Staphyloc 202 100.0 187 5 ABB81061 Canine lo  
130 25 100.0 4 ABBP38346 Staphyloc 203 100.0 187 5 ABB81061 Canine lo  
131 25 100.0 4 ABBP38346 Staphyloc 204 100.0 187 5 ABB81061 Canine lo  
132 25 100.0 4 ABBP38346 Staphyloc 205 100.0 187 5 ABB81061 Canine lo  
133 25 100.0 4 ABBP38346 Staphyloc 206 100.0 187 5 ABB81061 Canine lo  
134 25 100.0 4 ABBP38346 Staphyloc 207 100.0 187 5 ABB81061 Canine lo  
135 25 100.0 4 ABBP38346 Staphyloc 208 100.0 187 5 ABB81061 Canine lo  
136 25 100.0 4 ABBP38346 Staphyloc 209 100.0 187 5 ABB81061 Canine lo  
137 25 100.0 4 ABBP38346 Staphyloc 210 100.0 187 5 ABB81061 Canine lo  
138 25 100.0 4 ABBP38346 Staphyloc 211 100.0 187 5 ABB81061 Canine lo  
139 25 100.0 4 ABBP38346 Staphyloc 212 100.0 187 5 ABB81061 Canine lo  
140 25 100.0 4 ABBP38346 Staphyloc 213 100.0 187 5 ABB81061 Canine lo  
141 25 100.0 4 ABBP38346 Staphyloc 214 100.0 187 5 ABB81061 Canine lo  
142 25 100.0 4 ABBP38346 Staphyloc 215 100.0 187 5 ABB81061 Canine lo  
143 25 100.0 4 ABBP38346 Staphyloc 216 100.0 187 5 ABB81061 Canine lo  
144 25 100.0 4 ABBP38346 Staphyloc 217 100.0 187 5 ABB81061 Canine lo  
145 25 100.0 4 ABBP38346 Staphyloc 218 100.0 187 5 ABB81061 Canine lo  
146 25 100.0 4 ABBP38346 Staphyloc 219 100.0 187 5 ABB81061 Canine lo  
147 25 100.0 4 ABBP38346 Staphyloc 220 100.0 187 5 ABB81061 Canine lo  
148 25 100.0 4 ABBP38346 Staphyloc 221 100.0 187 5 ABB81061 Canine lo  
149 25 100.0 4 ABBP38346 Staphyloc 222 100.0 187 5 ABB81061 Canine lo  
150 25 100.0 4 ABBP38346 Staphyloc 223 100.0 187 5 ABB81061 Canine lo  
151 25 100.0 4 ABBP38346 Staphyloc 224 100.0 187 5 ABB81061 Canine lo  
152 25 100.0 4 ABBP38346 Staphyloc 225 100.0 187 5 ABB81061 Canine lo  
153 25 100.0 4 ABBP38346 Staphyloc 226 100.0 187 5 ABB81061 Canine lo  
154 25 100.0 4 ABBP38346 Staphyloc 227 100.0 187 5 ABB81061 Canine lo  
155 25 100.0 4 ABBP38346 Staphyloc 228 100.0 187 5 ABB81061 Canine lo  
156 25 100.0 4 ABBP38346 Staphyloc 229 100.0 187 5 ABB81061 Canine lo  
157 25 100.0 4 ABBP38346 Staphyloc 230 100.0 187 5 ABB81061 Canine lo  
158 25 100.0 4 ABBP38346 Staphyloc 231 100.0 187 5 ABB81061 Canine lo  
159 25 100.0 4 ABBP38346 Staphyloc 232 100.0 187 5 ABB81061 Canine lo  
160 25 100.0 4 ABBP38346 Staphyloc 233 100.0 187 5 ABB81061 Canine lo  
161 25 100.0 4 ABBP38346 Staphyloc 234 100.0 187 5 ABB81061 Canine lo  
162 25 100.0 4 ABBP38346 Staphyloc 235 100.0 187 5 ABB81061 Canine lo  
163 25 100.0 4 ABBP38346 Staphyloc 236 100.0 187 5 ABB81061 Canine lo  
164 25 100.0 4 ABBP38346 Staphyloc 237 100.0 187 5 ABB81061 Canine lo  
165 25 100.0 4 ABBP38346 Staphyloc 238 100.0 187 5 ABB81061 Canine lo  
166 25 100.0 4 ABBP38346 Staphyloc 239 100.0 187 5 ABB81061 Canine lo  
167 25 100.0 4 ABBP38346 Staphyloc 240 100.0 187 5 ABB81061 Canine lo  
168 25 100.0 4 ABBP38346 Staphyloc 241 100.0 187 5 ABB81061 Canine lo  
169 25 100.0 4 ABBP38346 Staphyloc 242 100.0 187 5 ABB81061 Canine lo  
170 25 100.0 4 ABBP38346 Staphyloc 243 100.0 187 5 ABB81061 Canine lo  
171 25 100.0 4 ABBP38346 Staphyloc 244 100.0 187 5 ABB81061 Canine lo

172 100.0 165 5 AAM48284 Pheromone  
173 100.0 166 4 ABG17237 Novel hum  
174 100.0 167 3 AAG18653 Arabidops  
175 100.0 167 4 AAU16185 Human nov  
176 100.0 167 6 ABU55254 Human nov  
177 100.0 168 3 AAG18652 Arabidops  
178 100.0 170 3 AAG58217 Arabidops  
179 100.0 171 3 AAG10427 Arabidops  
180 100.0 172 2 AAR12796 Human imm  
181 100.0 172 2 AAR33377 Soluble C  
182 100.0 172 2 ABU04233 Human exp  
183 100.0 172 6 ABU04233 Human exp  
184 100.0 172 6 ABU04233 Human exp  
185 100.0 175 3 AAG27379 Arabidops  
186 100.0 176 4 AAG33496 Arabidops  
187 100.0 177 3 AAG33496 Arabidops  
188 100.0 178 4 AAU16138 Human nov  
189 100.0 178 6 ABU55207 Mycoplasma  
190 100.0 179 6 ABM69668 Phototrab  
191 100.0 179 6 ABM69668 Phototrab  
192 100.0 179 6 ABM69668 Phototrab  
193 100.0 179 6 ABM69668 Phototrab  
194 100.0 183 5 ABP40491 Staphyloc  
195 100.0 183 5 ABP40491 Staphyloc  
196 100.0 187 5 ABB81061 Canine lo  
197 100.0 187 5 ABB81061 Canine lo  
198 100.0 187 5 ABB81061 Canine lo  
199 100.0 187 5 ABB81061 Canine lo  
200 100.0 187 5 ABB81061 Canine lo  
201 100.0 187 5 ABB81061 Canine lo  
202 100.0 187 5 ABB81061 Canine lo  
203 100.0 187 5 ABB81061 Canine lo  
204 100.0 187 5 ABB81061 Canine lo  
205 100.0 187 5 ABB81061 Canine lo  
206 100.0 187 5 ABB81061 Canine lo  
207 100.0 187 5 ABB81061 Canine lo  
208 100.0 187 5 ABB81061 Canine lo  
209 100.0 187 5 ABB81061 Canine lo  
210 100.0 187 5 ABB81061 Canine lo  
211 100.0 187 5 ABB81061 Canine lo  
212 100.0 187 5 ABB81061 Canine lo  
213 100.0 187 5 ABB81061 Canine lo  
214 100.0 187 5 ABB81061 Canine lo  
215 100.0 187 5 ABB81061 Canine lo  
216 100.0 187 5 ABB81061 Canine lo  
217 100.0 187 5 ABB81061 Canine lo  
218 100.0 187 5 ABB81061 Canine lo  
219 100.0 187 5 ABB81061 Canine lo  
220 100.0 187 5 ABB81061 Canine lo  
221 100.0 187 5 ABB81061 Canine lo  
222 100.0 187 5 ABB81061 Canine lo  
223 100.0 187 5 ABB81061 Canine lo  
224 100.0 187 5 ABB81061 Canine lo  
225 100.0 187 5 ABB81061 Canine lo  
226 100.0 187 5 ABB81061 Canine lo  
227 100.0 187 5 ABB81061 Canine lo  
228 100.0 187 5 ABB81061 Canine lo  
229 100.0 187 5 ABB81061 Canine lo  
230 100.0 187 5 ABB81061 Canine lo  
231 100.0 187 5 ABB81061 Canine lo  
232 100.0 187 5 ABB81061 Canine lo  
233 100.0 187 5 ABB81061 Canine lo  
234 100.0 187 5 ABB81061 Canine lo  
235 100.0 187 5 ABB81061 Canine lo  
236 100.0 187 5 ABB81061 Canine lo  
237 100.0 187 5 ABB81061 Canine lo  
238 100.0 187 5 ABB81061 Canine lo  
239 100.0 187 5 ABB81061 Canine lo  
240 100.0 187 5 ABB81061 Canine lo  
241 100.0 187 5 ABB81061 Canine lo  
242 100.0 187 5 ABB81061 Canine lo  
243 100.0 187 5 ABB81061 Canine lo  
244 100.0 187 5 ABB81061 Canine lo

Aag35079 Arabidops  
Abg03470 Novel hum  
Aag16530 Novel hum  
Aag33239 Novel hum  
Aag34872 Arabidops  
Aag07710 Arabidops  
Abb81058 Canine lo  
Aag81564 S. epider  
Abu42795 Protein e  
Aay32792 Growth ho  
Aabp38346 Staphyloc  
Aag54766 Arabidops  
Abb48396 Listeria  
Aap98021 Amino aci  
Abb81059 Canine lo  
Aag79296 Amino aci  
Abu58225 Rice stre  
Aap90474 Human int  
Aar03204 Human int  
Aar05289 Human int  
Aar05288 Human int  
Abb81060 Canine lo  
Aag54969 Arabidops  
Aag58966 Arabidops  
Abg11951 Novel hum  
Aar15431 Interleuk  
Aar21695 Delta(1-1  
Aap90186 New inter  
Aap70415 Sequence  
Aar90185 Human mat  
Aay96876 Human int  
Aac06657 Human int  
Abu58224 Corn stre  
Aar26985 Human IL-  
Abb70019 Drosophil  
Aar06267 Interleuk  
Aag34678 Arabidops  
Aag19271 Human G-P  
Aau27474 Human G-F  
Aag66509 RGS prote  
Abg25576 Novel hum  
Abm73654 DNA clone  
Ade31085 Human dia  
Aay17994 Human int  
Aar26984 Human IL-  
Aap90360 Recombina  
Aag27380 Arabidops  
Abu08122 Mature ra  
Abb06784 Human can  
Aap90472 Human int  
Aap70989 Sequence  
Aap82006 Part of I  
Aar03007 Interleuk  
Aar15429 Asp(36),  
Aar21692 (16G, 36D  
Aar21693 (36D, 141  
Aar33543 Human IL-  
Aar5364 Interleuk  
Aar90187 New inter  
Aar90182 Interleuk  
Aar90184 New inter  
Aaw31667 Leaderles  
Aag34677 Arabidops  
Abg71738 Human int  
Apu08118 Mature hu  
Aap82077 Recombina  
Aar23003 Human IL-  
Abb61412 Drosophil  
Aag31586 Streptoco  
Aay42392 A novel P  
Abb70935 Drosophil  
Adc96887 E. faeciu

Aam48284 Pheromone  
Abg17237 Novel hum  
Aag18653 Arabidops  
Aau16185 Human nov  
Abu55254 Human nov  
Aag18652 Arabidops  
Aag58217 Arabidops  
Aag10427 Arabidops  
Aar12796 Human imm  
Aar33377 Soluble C  
Aa47199 CD23. 8/1  
Abu04232 Human exp  
Abu04233 Human exp  
Abu04233 Human exp  
Aag27379 Arabidops  
Aag21258 Novel hum  
Aag33496 Arabidops  
Aau16138 Human nov  
Abu55207 Human nov  
Aay50056 Mycoplasma  
Abm69668 Phototrab  
Abm69668 Phototrab  
Abp40491 Staphyloc  
Abb81061 Canine lo  
Abu23987 Protein e  
Aag54929 Arabidops  
Aag33495 Arabidops  
Abu42271 Protein e  
Aag12209 Arabidops  
Aag43176 Arabidops  
Aag82988 S. epider  
Aag46862 Arabidops  
Abu39867 Protein e  
Aae08929 Rice subm  
Aab40714 Staphyloc  
Abu36151 Protein e  
Abb47473 Listeria  
Abu32939 Protein e  
Abu41536 Protein e  
Abu33585 Protein e  
Abu15567 Protein e  
Aay77144 Botulinum  
Aau23486 Novel hum  
Abg05044 Novel hum  
Abu24800 Protein e  
Abb1062 Canine lo  
Aay28793 Schizosac  
Aag2201 S. epider  
Abj18781 Pseudomon  
Aau03589 Human DNA  
Aag81294 Human AFP  
Abp69783 Human pol  
Abu19341 Protein e  
Abu18752 Protein e  
Abg91486 Purine/py  
Abg91485 Purine/py  
Aay36883 Amino aci  
Aae3602 Trypanoso  
Aae3601 Leishmani  
Aae3599 Trypanoso  
Ada15477 A. thalia  
Add55718 Thalecres  
Aam83850 Human imm  
Aap57762 Human pho  
Abu20089 Protein e  
Ade00077 Enterohae  
Aag81790 S. epider  
Abb50019 Listeria  
Aag33494 Arabidops  
Aag58216 Arabidops  
Aap82576 Water sol  
Aap81175 Sequence

## ALIGNMENTS

OS Synthetic.  
 XX EP943624-A1.  
 XX 22-SEP-1999.  
 XX 12-MAR-1998; 98EP-00200799.  
 XX 12-MAR-1998; 98EP-00200799.  
 XX (UYUT-) RIJKSUNIV UTRECHT.  
 XX WPI; 1999-510568/43.  
 XX Controlling the availability and/or signal transduction capability of a  
 XX cell surface receptor, useful for treating growth hormone deficiencies.  
 XX Disclosure; Page 9; 36pp; English.  
 XX This sequence is a polypeptide from the immunoglobulin E receptor.  
 XX Sequences (AA32794-Y32823) are examples of polypeptide sequences found  
 XX at or near the ubiquitin/proteasome complex binding site located on the  
 XX intracellular part of a cell surface receptor. These sequences are used  
 XX in a method for controlling the availability and signal transduction  
 XX capability of a cell surface receptor by administering an inhibitor that  
 XX is capable of inhibiting proteolytic cleavage of the receptor. Inhibition  
 XX of this proteolytic cleavage results in the receptors being present on  
 XX the surface for longer and therefore signalling for longer to the  
 XX interior of the cell. This increases the sensitivity of cells to any  
 XX hormones which might be present. The inhibitor is either derived from,  
 XX competes with or binds to a polypeptide sequence of which sequences  
 XX AA32794-Y32823 are examples. The inhibitor may be used to treat muscle  
 XX wasting, associated with disorders such as renal tubular defects, AIDS,  
 XX uraemia, diabetes, Cushing's syndrome, cachexias, eating disorders, and  
 XX after stress and during neuromuscular disease  
 XX Sequence 12 AA;  
 XX Query Match 100.0%; Score 25; DB 2; Length 12;  
 XX Best Local Similarity 50.0%; Pred. No. 1.8e+02;  
 XX Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEFIXDX 8  
 DB :|||:::  
 5 GEFIWVDG 12  
 RESULT 3  
 AAY32793  
 ID AAY32793 standard; peptide; 12 AA.  
 XX AAY32793;  
 XX AAY32793;  
 XX 09-NOV-1999 (first entry)  
 XX Growth hormone receptor derived polypeptide.  
 XX Signal transduction; proteolytic cleavage; growth hormone receptor;  
 XX proteasome binding site; muscle wasting; renal tubular defect; uraemia;  
 XX diabetes; Cushing's disease; eating disorder; AIDS; ubiquitin;  
 XX growth hormone deficiency.  
 XX Mammalia.  
 XX EP943624-A1.  
 XX 22-SEP-1999.  
 XX 12-MAR-1998; 98EP-00200799.  
 XX 12-MAR-1998; 98EP-00200799.  
 XX (UYUT-) RIJKSUNIV UTRECHT.

XX WPI; 1999-510568/43.  
 XX Controlling the availability and/or signal transduction capability of a  
 XX cell surface receptor, useful for treating growth hormone deficiencies.  
 XX Disclosure; Page 5; 36pp; English.  
 XX This sequence is derived from the growth hormone receptor polypeptide  
 XX AA32792. This sequence and variants (AA32794-Y32823) of it are examples  
 XX of polypeptide sequences found at or near the ubiquitin/proteasome  
 XX binding site located on the intracellular part of a cell surface  
 XX receptor. These sequences are used in a method for controlling the  
 XX availability and signal transduction capability of a cell surface  
 XX receptor by administering an inhibitor that is capable of inhibiting  
 XX proteolytic cleavage of the receptor. Inhibition of this proteolytic  
 XX cleavage results in the receptors being present on the surface for longer  
 XX and therefore signalling for longer to the interior of the cell. This  
 XX increases the sensitivity of cells to any hormones which might be  
 XX present. The inhibitor is either derived from, competes with or binds to  
 XX a polypeptide sequence of which sequences AA32794-Y32823 are examples.  
 XX The inhibitor may be used to treat muscle wasting, associated with  
 XX disorders such as renal tubular defects, uraemia, diabetes, Cushing's  
 XX syndrome, cachexias, eating disorders, AIDS, after stress and during  
 XX neuromuscular disease  
 XX Sequence 12 AA;  
 XX Query Match 100.0%; Score 25; DB 2; Length 12;  
 XX Best Local Similarity 50.0%; Pred. No. 1.8e+02;  
 XX Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEFIXDX 8  
 DB :|||:::  
 3 VEFIELDI 10  
 RESULT 4  
 AAY32789  
 ID AAY32789 standard; peptide; 12 AA.  
 XX AAY32789;  
 XX AAY32789;  
 XX 09-NOV-1999 (first entry)  
 XX Growth hormone receptor polypeptide for inhibition of receptor cleavage.  
 XX Growth hormone receptor; signal transduction; proteolytic cleavage;  
 XX ubiquitin; proteasome binding site; muscle wasting; renal tubular defect;  
 XX uraemia; diabetes; Cushing's disease; eating disorder; AIDS.  
 XX Mammalia.  
 XX EP943624-A1.  
 XX 22-SEP-1999.  
 XX 12-MAR-1998; 98EP-00200799.  
 XX 12-MAR-1998; 98EP-00200799.  
 XX (UYUT-) RIJKSUNIV UTRECHT.  
 XX WPI; 1999-510568/43.  
 XX Controlling the availability and/or signal transduction capability of a  
 XX cell surface receptor, useful for treating growth hormone deficiencies.  
 XX Claim 11; Page 27; 36pp; English.  
 XX This sequence is from the intracellular domain of the growth hormone  
 XX receptor. This sequence is an example of a binding site motif AAY32790.  
 XX This motif is used in a method for controlling the availability and



CC signal transduction capability of a cell surface receptor by  
 CC administering an inhibitor that is capable of inhibiting proteolytic  
 CC cleavage of the receptor. Inhibition of this proteolytic cleavage results  
 CC in the receptors being present on the surface for longer and therefore  
 CC signalling for longer to the interior of the cell. This increases the  
 CC sensitivity of cells to any hormones which might be present. The binding  
 CC motif, of which this sequence is an example, is located at or around a  
 CC ubiquitin and/or ubiquitin/proteasome complex binding site. The inhibitor  
 CC of the method is either derived from, competes with, or binds to this  
 CC binding motif. The inhibitor may be used to treat muscle wasting,  
 CC associated with disorders such as renal tubular defects, uraemia,  
 CC diabetes, Cushing's syndrome, cachexias, eating disorders, AIDS, after  
 CC stress and during neuromuscular disease  
 XX  
 SQ Sequence 12 AA;

Query Match 100.0%; Score 25; DB 2; Length 12;  
 Best Local Similarity 50.0%; Pred. No. 1.8e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFTXDXD 8  
 :|||:|:  
 Db 5 VEFIELDI 12

RESULT 5  
 AAY32794  
 ID AAY32794 standard; peptide; 12 AA.

XX AC AAY32794;

XX DT 09-NOV-1999 (first entry)

XX DE Chicken growth hormone receptor derived polypeptide.

XX KW Signal transduction; proteolytic cleavage; growth hormone receptor;  
 KW proteasome binding site; muscle wasting; renal tubular defect; uraemia;  
 KW diabetes; Cushing's disease; eating disorder; AIDS; ubiquitin;  
 KW growth hormone deficiency; inhibitor.

XX OS Gallus gallus.

XX PN EP943624-AL.

XX PD 22-SEP-1999.

XX PF 12-MAR-1998; 98EP-00200799.

XX PR 12-MAR-1998; 98EP-00200799.

XX PA (UYUT-) RIJKSUNIV UTRECHT.

XX WPI; 1999-510568/43.

XX PT Controlling the availability and/or signal transduction capability of a  
 XX cell surface receptor, useful for treating growth hormone deficiencies.

XX PS Disclosure; Page 9; 36pp; English.

XX CC This sequence is a polypeptide from the chicken growth hormone receptor.  
 CC Sequences (AAY32794-Y32823) are examples of polypeptide sequences found  
 CC at or near the ubiquitin/proteasome complex binding site located on the  
 CC intracellular part of a cell surface receptor. These sequences are used  
 CC in a method for controlling the availability and signal transduction  
 CC capability of a cell surface receptor by administering an inhibitor that  
 CC is capable of inhibiting proteolytic cleavage of the receptor. Inhibition  
 CC of this proteolytic cleavage results in the receptors being present on  
 CC the surface for longer and therefore signalling for longer to the  
 CC interior of the cell. This increases the sensitivity of cells to any  
 CC hormones which might be present. The inhibitor is either derived from,  
 CC competes with or binds to a polypeptide sequence of which sequences  
 CC AAY32794-Y32823 are examples. The inhibitor may be used to treat muscle  
 CC wasting, associated with disorders such as renal tubular defects,

CC uraemia, diabetes, Cushing's syndrome, cachexias, eating disorders, AIDS,  
 CC after stress and during neuromuscular disease  
 XX  
 SQ Sequence 12 AA;

Query Match 100.0%; Score 25; DB 2; Length 12;  
 Best Local Similarity 50.0%; Pred. No. 1.8e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFTXDXD 8  
 :|||:|:  
 Db 5 VEFIELDI 12

RESULT 6  
 ABU03293

ID ABU03293 standard; protein; 13 AA.

XX AC ABU03293;

XX DT 29-JAN-2003 (first entry)

XX DE Human expressed protein tag (EPT) #73.

XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
 KW protease; protease inhibitor; transporter; cytoskeletal protein;  
 KW receptor; transcription factor; cancer; MHC;  
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX OS Homo sapiens.

XX PN WO200278524-A2.

XX PD 10-OCT-2002.

XX PF 28-MAR-2002; 2002WO-US009671.

XX PR 28-MAR-2001; 2001US-0279495P.

XX PR 21-MAY-2001; 2001US-0292544P.

XX PR 08-AUG-2001; 2001US-0310801P.

XX PR 01-OCT-2001; 2001US-0326370P.

XX PR 04-DEC-2001; 2001US-0336780P.

XX PR 20-FEB-2002; 2002US-0358985P.

XX PA (ZYCO-) ZYCOS INC.

XX PI Chicx RM, Tomlinson AJ, Urban RG;

XX WPI; 2003-040607/03.

XX PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 XX cytoskeletal proteins, receptors or transcription factors), useful for  
 XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 XX leukemia.

XX PS Claim 10; SEQ ID NO 73; 134pp; English.

XX CC The invention describes a purified polypeptide, which comprises a  
 XX fragment of a kinase, phosphatase, protease, protease inhibitor,  
 XX transporter, cytoskeletal protein, receptor or transcription factor. The  
 XX polypeptide is useful as an immunogenic composition for eliciting in a  
 XX mammal an immunogenic response directed against any of the purified  
 XX polypeptide. The purified polypeptide, or the antibody that binds to this  
 XX polypeptide, is useful for treating cancer. The polypeptide is also  
 XX useful for identifying compounds that binds to a naturally processed  
 XX class I or class II MHC-binding polypeptide. The polypeptides and  
 XX polynucleotides are particularly useful for treating or preventing  
 XX myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 XX lymphoma or leukaemia. These are also useful for screening agents for  
 XX treating the above mentioned diseases. This sequence represents an  
 XX expressed protein tag (EPT) isolated from human tissue for translational  
 XX profiling. Note: This sequence does not appear in the printed

CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 13 AA;

Query Match 100.0%; Score 25; DB 6; Length 13;  
 Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
 :|||:::  
 Db 2 GEFIVDQ 9

## RESULT 7

AAW33004  
 ID AAW33004 standard; peptide; 16 AA.

XX AC AAW33004;

XX DT 22-MAY-1998 (first entry)

XX DE Murine DP-3 protein E region nuclear localisation signal.

XX KW Assaying; cell cycle regulator; E region; DP-3 protein;

XX KW nuclear localisation signal; NLS; mouse; murine.

XX OS Mus sp.

XX XX WO9743647-A1.

XX PN 20-NOV-1997.

XX PF 15-MAY-1997; 97WO-GB001324.

XX PR 15-MAY-1996; 96GB-00010195.

XX XX (MEDI-) MEDICAL RES COUNCIL.

XX PA La Thangue NB, De La Luna S;

XX PI WPI; 1998-009053/01.

XX DR Assays for identifying regulators of cell cycle progression - comprise

PT expressing a protein having a transcription factor nuclear localisation

PT signal and determining the degree of nuclear localisation.

XX PS Claim 2; Page 57; 65pp; English.

XX CC The present sequence was used in the development of a novel assay for a  
 CC putative regulator of cell cycle progression. The assay comprises  
 CC expressing in a cell a protein comprising the E region and sufficient C-  
 CC terminal residues of a DP-3 protein to provide a functional nuclear  
 CC localisation signal (NLS), or the NLS of E2F-1, and a marker for nuclear  
 CC localisation. Then the degree of nuclear localisation in the presence and  
 CC absence of the putative regulator is determined. Regulators identified  
 CC using the assay can be used as potential regulators of cell  
 CC proliferation, or as models for rational drug design. Regulation of the  
 CC NLS may lead to effects such as enhanced cell division, blocking of cell  
 CC cycle progression or apoptosis. The regulators may also be used to design  
 CC other candidate regulatory compounds

XX Sequence 16 AA;

Query Match 100.0%; Score 25; DB 2; Length 16;  
 Best Local Similarity 50.0%; Pred. No. 2.4e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
 :|||:::  
 Db 7 REFIDSDF 14

## RESULT 8

ABP42844

ID ABP42844 standard; protein; 36 AA.

XX AC ABP42844;

XX DT 22-AUG-2002 (first entry)

XX DE Human ovarian antigen HPCR90, SEQ ID NO:3976.

XX KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive.

XX OS Homo sapiens.

XX XX WO200200677-A1.

XX PD 03-JAN-2002.

XX PF 07-JUN-2001; 2001WO-US018569.

XX PR 07-JUN-2000; 2000US-0209467P.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX PI Birse CE, Rosen CA;

XX XX WPI; 2002-147878/19.

XX DR N-PSDB; ABQ55921.

XX PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
 PT cancer), immune disorders, cardiovascular disorders and neurological  
 PT diseases.

XX PS Claim 11; SEQ ID NO 3976; 2922pp; English.

XX CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovarian and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents a human ovarian antigen of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 36 AA;

Query Match 100.0%; Score 25; DB 5; Length 36;

Best Local Similarity 50.0%; Pred. No. 5.6e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8

:|||::|:

Db 13 QEFIDDDK 20

RESULT 9

AAAY11253  
ID AAY11253 standard; protein; 37 AA.

XX AC

AAAY11253;

XX 20-MAY-1999 (first entry)

XX Streptococcus pneumoniae protein sequence ID NO:363.

DE Streptococcus pneumoniae strain 0100993; vaccine; immune response;  
XX Streptococcal infection; pneumococcal.

OS Streptococcus pneumoniae.

XX WO9737026-A1.

XX 09-OCT-1997.

XX 01-APR-1997; 97WO-US005306.

XX 02-APR-1996; 96US-0014690P.

XX 22-AUG-1996; 96US-0025788P.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO, Stodola RK;

XX WPI; 1997-503111/46.

XX N-PSDB; AAX30836.

XX Nucleic acids encoding pneumococcal polypeptide(s) - useful in vaccines,  
PT drug screening, etc.

XX Claim 6; Page 279; 354pp; English.

XX AAX30724 to AAX30946 represent genomic DNA sequences isolated from

CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences  
CC encode the novel proteins given in AAY1114 to AAY11367. The proteins,  
CC isolated from Streptococcus pneumoniae, can be used in vaccines against  
CC streptococcal infections and in assays for identifying compounds that  
CC inhibit or activate the activity of the proteins. The antagonists can be  
CC used to treat an individual having need to inhibit a bacterial protein.  
CC Vectors expressing the proteins can be used to induce a protective immune  
CC response in mammals

XX Sequence 37 AA;

Query Match

Best Local Similarity 100.0%; Score 25; DB 2; Length 37;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8

:|||::|:

Db 13 NEFIADDD 20

RESULT 10

AAAY36319

ID AAY36319 standard; protein; 45 AA.

XX AC

AAAY36319;

DT 17-SEP-1999 (first entry)

DE Human secreted protein encoded by gene 96.

XX Human; secreted protein; cancer; tumour; developmental abnormality;  
KW foetal deficiency; blood disorder; immune system disorder; inflammation;  
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;  
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;  
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;  
KW digestive disorder; endocrine disorder; infection; AIDS.

XX Homo sapiens.

OS WO9931117-A1.

XX 24-JUN-1999.

XX 17-DEC-1998; 98WO-US027059.

XX 18-DEC-1997; 97US-0068006P.

XX 18-DEC-1997; 97US-0068007P.

XX 18-DEC-1997; 97US-0068008P.

XX 18-DEC-1997; 97US-0068053P.

XX 18-DEC-1997; 97US-0068054P.

XX 18-DEC-1997; 97US-0068057P.

XX 18-DEC-1997; 97US-0068064P.

XX 18-DEC-1997; 97US-0070923P.

XX 19-DEC-1997; 97US-0068169P.

XX 19-DEC-1997; 97US-0068365P.

XX 19-DEC-1997; 97US-0068367P.

XX 19-DEC-1997; 97US-0068368P.

XX 19-DEC-1997; 97US-0068369P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Moore PA, Ruben SM, Carter KC, Shi Y, Rosen CA, Soppet DR;

XX Kyaw H, Wei Y, Florence K, Duan RD, Florence C, Greene JM, Feng P;

XX Ferrie AM, Yu G, Janat F, Ni J;

XX WPI; 1999-418749/35.

XX N-PSDB; AAX38011.

XX New isolated human genes encoding secreted polypeptides.

XX Claim 11; Page 401; 537pp; English.

XX AAX97916 to AAX98029 represent 110 isolated human secreted protein genes.  
CC AAX36224 to AAX36727 represent the secreted proteins encoded by the 110  
CC human genes. The genes and their corresponding secreted polypeptides are  
CC useful for preventing, treating or ameliorating medical conditions, e.g.  
CC by protein or gene therapy. Also pathological conditions can be diagnosed  
CC by determining the amount of the new polypeptides in a sample or by  
CC determining the presence of mutations in the new genes. Specific uses are  
CC described for each of the 110 genes, based on which tissues they are most  
CC highly expressed in, and include developing products for the diagnosis or  
CC treatment of cancer, tumours, developmental abnormalities and foetal  
CC deficiencies, blood disorders, diseases of the immune system, autoimmune  
CC diseases, inflammation, allergies, Alzheimer's and cognitive disorders,  
CC schizophrenia, arthritis, asthma, psoriasis, sepsis, skin disorders,  
CC atherosclerosis, diabetes, cardiovascular disorders, kidney disorders,  
CC digestive/endocrine disorders, infections and AIDS. The polypeptides are  
CC also useful for identifying their binding partners. The sequences given  
CC in AAX97907 to AAX97915 and AAX36223 are used in the exemplification of  
CC the present invention

XX Sequence 45 AA;

Query Match

Best Local Similarity 100.0%; Score 25; DB 2; Length 45;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:;  
Db 25 IEFIAADF 32

RESULT 11  
ADAL1680  
ID ADAL1680 standard; protein; 45 AA.  
XX  
AC ADAL1680;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Human novel secreted protein, SEQ ID NO 208.  
XX  
KW cancer; inflammation; immune disorder; neurological disorder;  
KW blood clotting disorder; food additive; food preservative;  
KW storage capability; fat content; nutritional component; human;  
KW secreted protein.  
XX  
OS Homo sapiens.  
XX  
PN US2003055236-A1.  
XX  
PD 20-MAR-2003.  
XX  
PF 14-MAR-2002; 2002US-00097065.  
XX  
PR 18-DEC-1997; 97US-0068006P.  
PR 18-DEC-1997; 97US-0068007P.  
PR 18-DEC-1997; 97US-0068008P.  
PR 18-DEC-1997; 97US-0068009P.  
PR 18-DEC-1997; 97US-0068053P.  
PR 18-DEC-1997; 97US-0068054P.  
PR 18-DEC-1997; 97US-0068055P.  
PR 18-DEC-1997; 97US-0068056P.  
PR 18-DEC-1997; 97US-0068057P.  
PR 18-DEC-1997; 97US-0068058P.  
PR 18-DEC-1997; 97US-0068059P.  
PR 18-DEC-1997; 97US-0068169P.  
PR 19-DEC-1997; 97US-0068365P.  
PR 19-DEC-1997; 97US-0068367P.  
PR 19-DEC-1997; 97US-0068368P.  
PR 19-DEC-1997; 97US-0068369P.  
PR 17-DEC-1998; 98WO-US027059.  
PR 17-JUN-1999; 99US-00334595.  
XX  
(HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Moore PA, Ruben SM, Carter KC, Shi Y, Rosen CA, Soppet DR;  
PI Kyaw H, Wei Y, Florence KA, Duan DR, Florence C, Greene JM, Peng P;  
PI Ferrie AM, Yu G, Janat F, Ni J;  
XX  
DR WPI: 2003-567105/53.  
DR N-PSDB; ADA11556.  
XX  
PT New secreted HKABT24 nucleic acid molecules and polypeptides, useful for  
PT preventing, treating, or ameliorating a medical condition, such as  
PT cancer, inflammation, immune disorders, neurological and blood clotting  
PT disorders.  
XX  
PS Claim 11; SEQ ID NO 208; 118pp; English.  
XX  
CC The invention relates to an isolated HKABT24 nucleic acid molecule. The  
CC polypeptides, nucleic acids and antibodies are useful for diagnosing a  
CC pathological condition or a susceptibility to a pathological condition,  
CC for preventing, treating, or ameliorating a medical condition, such as  
CC cancer, inflammation and other immune disorders, neurological and blood  
CC clotting disorders. The nucleic acids are also useful for chromosome  
CC identification, radiation hybrid mapping or long-range restriction  
CC mapping. The polypeptides and antibodies are useful for providing  
CC immunohistological probes for differential identification of the tissues  
CC immunohistochemistry assays. The polypeptide, polynucleotide, agonist or  
CC antagonist may also be used as a food additive or preservative to

CC increase or decrease storage capabilities, fat content or other  
CC nutritional components. The present sequence represents the amino acid  
CC sequence of a novel human secreted protein. Note: The sequence data for  
CC this patent did not form part of the printed specification but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov.uk/sequence.html?DocID=20030055236.  
XX  
SQ Sequence 45 AA;  
Query Match 100.0%; Score 25; DB 6; Length 45;  
Best Local Similarity 50.0%; Pred. NO. 7e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:;  
Db 25 IEFIAADF 32

RESULT 12  
ABP72498  
ID ABP72498 standard; protein; 48 AA.  
XX  
AC ABP72498;  
XX  
DT 23-MAY-2003 (first entry)  
XX  
DE Phospholipid transporter consensus sequence.  
XX  
KW Phospholipid transporter; ATPase; enzyme; 67108.  
XX  
OS Mammalia.  
XX  
PN EPI266907-A1.  
XX  
PD 18-DEC-2002.  
XX  
PF 13-JUN-2002; 2002EP-00254113.  
XX  
PR 13-JUN-2001; 2001US-0297840P.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Curtis RAJ;  
XX  
DR WPI: 2003-203087/20.  
XX  
PT New 61708 nucleic acid molecules and polypeptides, useful for diagnosing  
PT or treating conditions related to aberrant activity or deficient  
PT phospholipid transporter function or expression, e.g. autoimmune disease,  
PT cancer or asthma.  
XX  
PS Disclosure; Page 66; 74pp; English.  
XX  
CC The present sequence is that of amino acids 7-54 of a phospholipid  
CC transporter consensus sequence, denoted ATPase hydrolase phosphorylation  
CC transmembrane ATP-binding magnesium 3.6.1-calcium transporting multigene.  
CC This sequence shows 56% identity with a phospholipid-transporting domain  
CC (amino acids 123-170) of a novel human phospholipid transporter of the  
CC invention, designated 67108 (see ABP72495). The invention provides 67108  
CC nucleic acids and polypeptides, vectors, host cells, antibodies and  
CC transgenic animals. The 61708 nucleic acids and polypeptides are useful  
CC for diagnosing, preventing or treating conditions related to aberrant  
CC activity or deficient phospholipid transporter function or expression,  
CC such as neurological disorders, cellular proliferative and/or  
CC differentiative disorders, immune disorders, inflammatory disorders,  
CC viral diseases, cardiovascular disorders, blood coagulation disorders,  
CC and hormonal disorders  
XX  
SQ Sequence 48 AA;  
Query Match 100.0%; Score 25; DB 6; Length 48;  
Best Local Similarity 50.0%; Pred. NO. 7.5e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY	1 XEFIXDX 8	DB	11 GEFIVWDG 18
DB	19 DEFIPADL 26		
RESULT 13			
ABG80735			
ID	ABG80735 standard; protein; 54 AA.		
XX			
AC	ABG80735;		
XX			
DT	29-NOV-2002 (first entry)		
XX			
DE	C-type lectin-like domain protein IGE-FCR.		
XX			
KW	Scaffold protein; C-type lectin-like domain; CTLD; alpha-helix;		
KW	beta-strand; connecting segment; 14loop region; tetranectin;		
KW	ligand-binding specificity; human; mouse; rat; chicken; trout; cow.		
XX			
OS	Unidentified.		
XX			
PN	WO200248189-A2.		
XX			
PD	20-JUN-2002.		
XX			
PF	13-DEC-2001; 2001WO-DK000825.		
XX			
PR	13-DEC-2000; 2000DK-00001872.		
PR	28-FEB-2001; 2001US-0272098P.		
XX			
FA	(BORE-) BOREAN PHARMA AS.		
XX			
PI	Etzerodt M, Holtet TL, Graversen NJH, Thogersen HC;		
DR	WFI; 2002-643278/69.		
XX			
PT	Protein comprising a variant of model C-type lectin-like domains (CTLD),		
PT	in which alpha helices, beta-strands, connecting segments are conserved		
PT	to maintain CTLD scaffold structure, while the loop region is altered.		
XX			
PS	Claim 3; Page 5; 168pp; English.		
XX			
CC	The present invention relates to a new protein with scaffold structure of		
CC	C-type lectin-like domains (CTLD). The invention comprises a variant of a		
CC	model CTLD where alpha-helices and beta-strands and connecting segments		
CC	are conserved such that scaffold structure of C-type lectin-like domains		
CC	(CTLD) is substantially maintained, while the 14loop region is altered by		
CC	amino acid substitution, deletion, insertion or their combination. The		
CC	invention is useful for preparing a library of nucleotide sequences		
CC	encoding related proteins by randomising part or all of the nucleic acid		
CC	sequence encoding the loop region of its CTLD. The artificial CTLD		
CC	protein products are preferable to antibody derivatives as each binding		
CC	site is a single structurally autonomous protein domain. When used as		
CC	components of compositions to be used for in vivo diagnostic or		
CC	therapeutic purposes, artificial CTLD protein products constructed on the		
CC	basis of human CTLDs are virtually identical to the corresponding natural		
CC	CTLD protein already present in the body and are therefore less		
CC	immunogenic to the patient. They also have a smaller size, and thus		
CC	provide tissue penetration and distribution, as well as shorter half life		
CC	in circulation. Since murine and human tetranectin are identical in		
CC	structure, straightforward swapping of polypeptide segments defining		
CC	ligand-binding specificity between murine and human tetranectin		
CC	derivatives may be achieved. The present amino acid sequence represents a		
CC	CTLD protein of the invention		
XX			
XX	Sequence 54 AA;		
QY	1 XEFIXDX 8		
XX			
Query Match	100.0%; Score 25; DB 5; Length 54;		
Best Local Similarity	50.0%; Pred. No. 8.5e+02;		
Matches	4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;		
DB	19 DEFIPADL 26		
RESULT 14			
AA116692			
ID	AA116692 standard; protein; 56 AA.		
XX			
AC	AA116692;		
XX			
DT	12-OCT-2001 (first entry)		
XX			
DE	Peptide #3126 encoded by probe for measuring cervical gene expression.		
XX			
KW	Probe; human; microarray; gene expression; cervical epithelial cell;		
KW	cervical cancer.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200157278-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	30-JAN-2001; 2001WO-US000670.		
XX			
PR	04-FEB-2000; 2000US-0180312P.		
PR	26-MAY-2000; 2000US-0207456P.		
PR	30-JUN-2000; 2000US-00608408.		
PR	03-AUG-2000; 2000US-00632366.		
PR	21-SEP-2000; 2000US-0234687P.		
PR	27-SEP-2000; 2000US-0236359P.		
PR	04-OCT-2000; 2000GB-00024263.		
XX			
PA	(MOLE-) MOLECULAR DYNAMICS INC.		
XX			
PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX	WFI; 2001-488901/53.		
DR			
XX	Human genome-derived single exon nucleic acid probes useful for analyzing		
PT	gene expression in human cervical epithelial cells.		
XX			
PS	Claim 27; SEQ ID NO 21518; 487pp; English.		
XX			
CC	The present invention relates to human single exon nucleic acid probes		
CC	(SENPs: see AA110068-AA128459). The present sequence is a peptide encoded		
CC	by one such probe. The SENPs are derived from human HeLa cells. The SENPs		
CC	can be used to produce a single exon microarray, which can be used for		
CC	measuring human gene expression in a sample derived from human cervical		
CC	epithelial cells. By measuring gene expression, the probes are therefore		
CC	useful in grading and/or staging of diseases of the cervix, notably		
CC	cervical cancer. Note: The sequence data for this patent did not form		
CC	part of the printed specification, but was obtained in electronic format		

```

XX DE Peptide #3183 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human foetal liver.
XX PS Claim 27; SEQ ID NO 28312; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX CC human gene expression in a sample derived from human foetal liver. The
XX CC single exon nucleic acid probes may be used for predicting, measuring and
XX CC displaying gene expression in samples derived from human foetal liver. The
XX CC present sequence is a peptide encoded by a single exon nucleic acid probe
XX CC of the invention. Note: The sequence data for this patent did not form
XX CC part of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 56 AA;

Query Match 100.0%; Score 25; DB 4; Length 56;
Best Local Similarity 50.0%; Pred. No. 8.8e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 37 QEFINLDG 44

RESULT 16
AAM29179
ID AAM29179 standard; protein; 56 AA.
XX AC AAM29179;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #3216 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.

```

```

PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX PS Claim 27; SEQ ID NO 29448; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX CC see AA131315-AA157546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders
XX SQ Sequence 56 AA;

Query Match 100.0%; Score 25; DB 4; Length 56;
Best Local Similarity 50.0%; Pred. No. 8.8e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 37 QEFINLDG 44

RESULT 17
ABB30514
ID ABB30514 standard; peptide; 56 AA.
XX AC ABB30514;
XX DT 01-FEB-2002 (first entry)
XX DE Peptide #3165 encoded by breast cell single exon nucleic acid probe.
XX KW Human; microarray; single exon probe; gene expression; breast; disease;
XX KW cancer.
XX OS Homo sapiens.
XX PN WO200157271-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000662.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX DR New spatially-addressable set of single exon nucleic acid probes, useful
XX PT

```

PT for measuring gene expression in sample derived from human breast,  
 PT comprises number of single exon nucleic acid probes.  
 XX  
 PS Claim 27; SEQ ID NO 13482; 327pp + Sequence Listing; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and BT 474 cells. The method involves contacting the  
 CC probes with a collection of detectably labelled nucleic acids derived  
 CC from mRNA of human breast, and then measuring the label bound to each  
 CC probe of the microarray. The probes are useful for verifying the  
 CC expression of regions of genomic DNA predicted to encode proteins. They  
 CC are useful for gene discovery, and for determining predisposition and/or  
 CC prognosing breast disease. Gene expression analysis is useful for  
 CC assessing the toxicity of chemical agents on cells. The microarray of  
 CC this invention presents a far greater diversity of probes for measuring  
 CC gene expression, with far less bias than expressed sequence tag  
 CC microarrays. The method is suitable for rapid production of functional  
 CC information from genomic sequence. The present sequence is a peptide  
 CC encoded by a single exon nucleic acid probe of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 56 AA;

Query Match 100.0%; Score 25; DB 4; Length 56;  
 Best Local Similarity 50.0%; Pred. No. 8.8e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||:|:  
 Db 37 QEFINDG 44

RESULT 18  
 ABB21106  
 ID ABB21106 standard; protein; 56 AA.  
 XX  
 AC ABB21106;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Protein #3105 encoded by probe for measuring heart cell gene expression.  
 XX  
 KW Human; gene expression; heart; microarray; vascular system;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157274-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000666.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-48899/53.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts.

XX  
 PS Claim 15; SEQ ID NO 22876; 530pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease,  
 CC hypertension, cardiac arrhythmias and congenital heart disease.  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 56 AA;

Query Match 100.0%; Score 25; DB 4; Length 56;  
 Best Local Similarity 50.0%; Pred. No. 8.8e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||:|:  
 Db 37 QEFINDG 44

RESULT 19  
 AAM68870  
 ID AAM68870 standard; protein; 56 AA.  
 XX  
 AC AAM68870;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 29176.  
 XX  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukaemia; lymphoma; myeloma.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157276-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000668.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488900/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human bone marrow.  
 XX  
 PS Example 4; SEQ ID NO 29176; 658pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention

```
XX SQ Sequence 56 AA;
Query Match 100.0%; Score 25; DB 4; Length 56;
Best Local Similarity 50.0%; Pred. No. 8.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 37 QEFINLDG 44

RESULT 20
AAM56491
ID AAM56491 standard; protein; 56 AA.
AC AAM56491;
XX
XX 05-NOV-2001 (first entry)
DT Human brain expressed single exon probe encoded protein SEQ ID NO: 28596.
DE Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
XX Homo sapiens.
OS
XX WO200157275-A2.
PN
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000667.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-48898/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
XX Claim 27; SEQ ID NO 29179; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABG47348-ABG59930 represent human
CC liver single exon encoded peptides of the invention. Note: The sequence
CC information for this patent does not appear in the printed specification
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 56 AA;
SQ
Query Match 100.0%; Score 25; DB 4; Length 56;
Best Local Similarity 50.0%; Pred. No. 8.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 37 QEFINLDG 44

RESULT 22
AAM04409
ID AAM04409 standard; protein; 56 AA.
AC AAM04409;
XX
XX 09-OCT-2001 (first entry)
DT Peptide #3091 encoded by probe for measuring breast gene expression.
DE
```



XX Probe; human; breast disease; breast cancer; development disorder;  
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 XX Homo sapiens.  
 OS WO200157270-A2.  
 PN 09-AUG-2001.  
 PD 29-JAN-2001; 2001WO-US000661.  
 XX 04-FEB-2000; 2000US-0180312P.  
 XX 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-476286/51.  
 XX Novel single exon nucleic acid probe used to measuring gene expression in  
 PT a human breast.  
 PT Claim 27; SEQ ID NO 13149; 322pp; English.  
 PS The present invention relates to novel single exon nucleic acid probes  
 CC (see AA100010-AA110067). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for measuring human gene expression in  
 CC a human breast sample, where the probe hybridises at high stringency to a  
 CC nucleic acid expressed in the human breast. The probes are useful for  
 CC predicting, diagnosing, grading, staging, monitoring and prognosing  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer, disorders of development,  
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
 CC breast disease and non-carcinoma tumours. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 56 AA;  
 SQ  
 Query Match 100.0%; Score 25; DB 4; Length 56;  
 Best Local Similarity 50.0%; Pred. No. 8.8e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEPIXXDX 8  
 Db :|||:|:  
 37 QEPINLDG 44  
 RESULT 23  
 ABG38451  
 ID ABG38451 standard; peptide; 56 AA.  
 XX AC ABG38451;  
 XX DT 19-AUG-2002 (first entry)  
 XX Human peptide encoded by genome-derived single exon probe SEQ ID 28116.  
 DE Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX Homo sapiens.  
 OS WO200186003-A2.  
 PN 15-NOV-2001.  
 PD 30-JAN-2001; 2001WO-US000665.  
 XX 04-FEB-2000; 2000US-0180312P.  
 XX 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples.  
 PT Claim 27; SEQ ID NO 28116; 634pp; English.  
 PS The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of probes  
 CC; the novel set of probes which hybridise at high stringency to a nucleic  
 CC acid expressed in the human lung; measuring gene expression in a sample  
 CC derived from human lung, comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of the  
 CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC algorithmically predicting at least one exon from genomic sequences of  
 CC the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene.  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC analysis, and for identifying exons in a gene, particularly using human  
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC present sequence is a peptide/protein encoded by a single exon probe of  
 CC the invention. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 56 AA;  
 SQ

Query Match 100.0%; Score 25; DB 5; Length 56;  
 Best Local Similarity 50.0%; Pred. No. 8.8e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:  
Db 37 QEFINLDG 44

## RESULT 24

AAAG02282  
ID AAG02282 standard; protein; 57 AA.

XX AC AAG02282;

XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein, SEQ ID NO: 6363.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX KW gene therapy; chromosome mapping.

XX OS Homo sapiens.

XX PN EPI033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-00200610.

XX PR 26-FEB-1999; 99US-0122487P.

XX PA (GSEST ) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX DR WPI; 2000-500381/45.

XX DR N-PSDB; AAC02288.

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX PS Claim 13; SEQ ID NO 6363; 71pp + Sequence Listing; English.

XX CC The present sequence is a polypeptide encoded by one of a large number of  
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different  
CC tissues. EST sequences usually correspond mainly to the 3' untranslated  
CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
CC sequences derived from the 5' ends of mRNAs and even in those cases where  
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
CC are also used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. They are used to obtain upstream regulatory sequences  
CC and to design expression and secretion vectors

XX SQ Sequence 57 AA;

Query Match 100.0%; Score 25; DB 3; Length 57;

Best Local Similarity 50.0%; Pred. No. 8.9e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:  
Db 39 FEFILIDF 46

## RESULT 25

AAAG10476  
ID AAG10476 standard; protein; 57 AA.

XX AC AAG10476;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 8815.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EPI033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

XX PR 05-MAR-1999; 99US-0123180P.

XX PR 09-MAR-1999; 99US-0123548P.

XX PR 23-MAR-1999; 99US-0125788P.

XX PR 25-MAR-1999; 99US-0126264P.

XX PR 29-MAR-1999; 99US-0126785P.

XX PR 01-APR-1999; 99US-0127462P.

XX PR 06-APR-1999; 99US-0128234P.

XX PR 16-APR-1999; 99US-0128714P.

XX PR 19-APR-1999; 99US-0130077P.

XX PR 21-APR-1999; 99US-0130449P.

XX PR 23-APR-1999; 99US-0130510P.

XX PR 28-APR-1999; 99US-0130891P.

XX PR 30-APR-1999; 99US-0131449P.

XX PR 30-APR-1999; 99US-0132048P.

XX PR 04-MAY-1999; 99US-0132407P.

XX PR 06-MAY-1999; 99US-0132484P.

XX PR 06-MAY-1999; 99US-0132485P.

XX PR 07-MAY-1999; 99US-0132486P.

XX PR 11-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

XX PR 14-MAY-1999; 99US-0132487P.

PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140335P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142330P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144864P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-01451145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.

PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155559P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160880P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 57;  
Best Local Similarity 50.0%; Pred. No. 8.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
Db 31 IEFICLDL 38

RESULT 26  
AAG24529  
ID AAG24529 standard; protein; 57 AA.  
XX  
AC AAG24529;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SQ ID NO: 28238.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.  
XX Arabidopsis thaliana.  
OS EP1033405-A2.  
XX 06-SEP-2000.  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130694P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 18-MAY-1999; 99US-0134370P.  
PR 19-MAY-1999; 99US-0134768P.  
PR 20-MAY-1999; 99US-0134941P.  
PR 21-MAY-1999; 99US-0135124P.  
PR 24-MAY-1999; 99US-0135353P.  
PR 25-MAY-1999; 99US-0136229P.  
PR 27-MAY-1999; 99US-0136021P.  
PR 28-MAY-1999; 99US-0136392P.  
PR 01-JUN-1999; 99US-0136782P.  
PR 03-JUN-1999; 99US-0137222P.  
PR 04-JUN-1999; 99US-0137528P.  
PR 07-JUN-1999; 99US-0137502P.  
PR 08-JUN-1999; 99US-0137724P.  
PR 10-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 14-JUN-1999; 99US-0138847P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 24-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 23-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.

PR 10-SEP-1999; 99US-0153070P.  
 PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154039P.  
 PR 20-SEP-1999; 99US-0154779P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 24-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155659P.  
 PR 28-SEP-1999; 99US-0156458P.  
 PR 29-SEP-1999; 99US-0156596P.  
 PR 04-OCT-1999; 99US-0157117P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157865P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 18-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160880P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161933P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 57;  
 Best Local Similarity 50.0%; Pred. No. 8.9e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Oy 1 XEFIXDX 8  
 Db 31 IEFICLDL 38

RESULT 27

ID AAW99092 standard; protein; 58 AA.

XX AC AAW99092;

DT 27-AUG-2003 (revised)

DT 04-JUN-1999 (first entry)

XX Oropouche NP protein fragment from a recombinant plasmid.

DE Oropouche virus; NP protein; Bunyavirus; Bunyaviridae; nucleocapsid;  
 KW immunodetection; seroprevalence; infection; vaccine.

XX Synthetic.  
 OS Unidentified.

XX WO9903875-A1.

PN

PD 28-JAN-1999.  
 XX 17-JUL-1998; 98WO-US014887.  
 XX 17-JUL-1997; 97US-0052848P.  
 XX (TEXA ) UNIV TEXAS SYSTEM.  
 XX Barrett ADT, Saeed WF, Tesh RB, Shope RE, Wang H;  
 XX WPI; 1999-132152/11.  
 XX N-PSDB; AAX19099.  
 PT New DNA encoding the nucleocapsid protein of Oropouche virus - used for  
 PT diagnosing Oropouche virus infection and in vaccines against this  
 PT infection.  
 XX Example 7; Fig 7; 60pp; English.  
 PS The present invention describes an Oropouche NP (nucleocapsid) protein.  
 CC The Oropouche virus belongs to the genus Bunyavirus of family  
 CC Bunyaviridae. The Oropouche NP protein can be used to detect anti-NP  
 CC antibodies, for diagnosis of Oropouche virus infection and for  
 CC determining seroprevalence. The nucleic acid sequence encoding the NP  
 CC protein is used to design: (i) oligonucleotide primers for diagnostic  
 CC reverse transcription polymerase chain reaction (RT-PCR) of viral RNA in  
 CC infected cell cultures or serum samples; or (ii) probes for hybridisation  
 CC reactions with viral RNA. Also, the nucleic acid sequence may be used in  
 CC vaccines to protect against Oropouche virus infection. The present  
 CC invention represents an Oropouche NP protein fragment from a recombinant  
 CC plasmid, which is used in an example from the present invention. (Updated  
 CC on 27-AUG-2003 to correct OS field.)  
 XX SQ Sequence 58 AA;

Query Match 100.0%; Score 25; DB 2; Length 58;  
 Best Local Similarity 50.0%; Pred. No. 9.1e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Oy 1 XEFIXDX 8  
 Db 41 SEFIENDV 48

RESULT 28

ABB43221  
 ID ABB43221 standard; peptide; 60 AA.

XX AC ABB43221;

XX 04-FEB-2002 (first entry)

XX Peptide #10727 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000669.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA

```
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
XX Claim 27; SEQ ID NO 35856; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a peptide encoded by a single exon nucleic acid probe
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 60 AA;
SQ
Query Match 100.0%; Score 25; DB 4; Length 60;
Best Local Similarity 50.0%; Pred. No. 9.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXXDX 8
DB 37 TEFINMDG 44
RESULT 29
AAM37061
ID AAM37061 standard; protein; 60 AA.
XX AC AAM37061;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #11098 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX FN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 27; SEQ ID NO 37330; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENPs;
XX see AA131315-AA157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
```

```
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders
XX
SQ Sequence 60 AA;
Query Match 100.0%; Score 25; DB 4; Length 60;
Best Local Similarity 50.0%; Pred. No. 9.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXXDX 8
DB 37 TEFINMDG 44
RESULT 30
AAM76953
ID AAM76953 standard; protein; 60 AA.
XX AC AAM76953;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 37259.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX FN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX Example 4; SEQ ID NO 37259; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX
SQ Sequence 60 AA;
Query Match 100.0%; Score 25; DB 4; Length 60;
Best Local Similarity 50.0%; Pred. No. 9.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXXDX 8
DB 37 TEFINMDG 44
```

```
RESULT 31
AAM64128
ID AAM64128 standard; protein; 60 AA.
XX
XX AAM64128;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 36233.
DE
XX
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
KW
XX
XX Homo sapiens.
OS
XX
XX W0200157275-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000667.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-488898/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
PT
XX
XX Claim 27; SEQ ID NO 37263; 658pp; English.
PS
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABG47348-ABG59930 represent human
CC liver single exon encoded peptides of the invention. Note: The sequence
CC information for this patent does not appear in the printed specification
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 60 AA;
SQ
Query Match 100.0%; Score 25; DB 4; Length 60;
Best Local Similarity 50.0%; Pred. No. 9.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db :|||:::
37 TEFINWDG 44
RESULT 32
ABG58615
ID ABG58615 standard; peptide; 60 AA.
XX
XX ABG58615;
AC
XX
XX 25-FEB-2003 (first entry)
DT
XX
XX Human liver peptide, SEQ ID No 37263.
DE
XX
XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW hypercholesterolaemia; coronary heart disease.
KW
XX
XX Homo sapiens.
OS
```

---

```
XX
PN
XX
XX W0200157273-A2.
PD
XX
XX 09-AUG-2001.
PF
XX
XX 30-JAN-2001; 2001WO-US000664.
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-488898/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
PT
XX
XX Claim 27; SEQ ID NO 37263; 658pp; English.
PS
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABG47348-ABG59930 represent human
CC liver single exon encoded peptides of the invention. Note: The sequence
CC information for this patent does not appear in the printed specification
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 60 AA;
SQ
Query Match 100.0%; Score 25; DB 4; Length 60;
Best Local Similarity 50.0%; Pred. No. 9.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db :|||:::
37 TEFINWDG 44
RESULT 33
ABG46060
ID ABG46060 standard; peptide; 60 AA.
XX
XX ABG46060;
AC
XX
XX 19-AUG-2002 (first entry)
DT
XX
XX Human peptide encoded by genome-derived single exon probe SEQ ID 35725.
DE
XX
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
XX Homo sapiens.
OS
```

XX WO200186003-A2.  
XX 15-NOV-2001.  
XX 30-JAN-2001; 2001WO-US000665.  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
XX measure gene expression in human lung samples.  
XX Claim 27; SEQ ID NO 35725; 634pp; English.  
XX The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human lung comprising single exon nucleic acid probes having one of  
XX 12614 nucleic acid sequences mentioned in the specification, or their  
XX complements or the 12387 open reading frames derived from the 12614  
XX probes. Also included are a microarray comprising the novel set of probes  
XX; the novel set of probes which hybridise at high stringency to a nucleic  
XX acid expressed in the human lung; measuring gene expression in a sample  
XX derived from human lung, comprising (a) contacting the array with a  
XX collection of detectably labeled nucleic acids derived from human lung  
XX mRNA, and (b) measuring the label detectably bound to each probe of the  
XX array; identifying exons in a eukaryotic genome, comprising (a)  
XX algorithmically predicting at least one exon from genomic sequences of  
XX the eukaryote; and (b) detecting specific hybridisation of detectably  
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
XX having a fragment identical to the predicted exon, the probe is included  
XX in the above mentioned microarray; assigning exons to a single gene,  
XX comprising (a) identifying exons from genomic sequence by the method  
XX above and (b) measuring the expression of each of the exons in several  
XX tissues and/or cell types using hybridisation to a single exon  
XX microarrays having a probe with the exon, where a common pattern of  
XX expression of the exons in the tissues and/or cell types indicates that  
XX the exons should be assigned to a single gene; a peptide comprising one  
XX of 12011 sequences, mentioned in the specification, or encoded by the  
XX probes/open reading frames (ORF). The probes are used for gene expression  
XX analysis, and for identifying exons in a gene, particularly using human  
XX lung derived mRNA and for the study of lung diseases such as asthma, lung  
XX cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
XX disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
XX tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
XX Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
XX histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
XX Karagenen syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
XX dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
XX present sequence is a peptide/protein encoded by a single exon probe of  
XX the invention. Note: The sequence data for this patent did not form part  
XX of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 60 AA;  
XX  
XX Query Match 100.0%; Score 25; DB 5; Length 60;  
XX Best Local Similarity 50.0%; Pred. No. 9.4e+02;  
XX Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
XX 1 XEFTXXDX 8  
XX :|||:-|

Db 37 TEFINMDG 44  
RESULT 34  
AAG34535  
ID AAG34535 standard; protein; 61 AA.  
XX AC AAG34535;  
XX 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 42037.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX Arabidopsis thaliana.  
XX EPI033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 29-MAR-1999; 99US-0126264P.  
XX 01-APR-1999; 99US-0126785P.  
XX 06-APR-1999; 99US-0127462P.  
XX 08-APR-1999; 99US-0128234P.  
XX 16-APR-1999; 99US-0128714P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 28-APR-1999; 99US-0130891P.  
XX 30-APR-1999; 99US-0131449P.  
XX 04-MAY-1999; 99US-0132048P.  
XX 05-MAY-1999; 99US-0132407P.  
XX 06-MAY-1999; 99US-0132484P.  
XX 06-MAY-1999; 99US-0132485P.  
XX 07-MAY-1999; 99US-0132486P.  
XX 11-MAY-1999; 99US-0132863P.  
XX 14-MAY-1999; 99US-0134256P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134221P.  
XX 18-MAY-1999; 99US-0134370P.  
XX 19-MAY-1999; 99US-0134786P.  
XX 20-MAY-1999; 99US-0134941P.  
XX 21-MAY-1999; 99US-0135124P.  
XX 24-MAY-1999; 99US-0135353P.  
XX 25-MAY-1999; 99US-0135629P.  
XX 27-MAY-1999; 99US-0136021P.  
XX 28-MAY-1999; 99US-0136382P.  
XX 01-JUN-1999; 99US-0136782P.  
XX 03-JUN-1999; 99US-0137222P.  
XX 04-JUN-1999; 99US-0137528P.  
XX 07-JUN-1999; 99US-0137502P.  
XX 08-JUN-1999; 99US-0137724P.  
XX 10-JUN-1999; 99US-0138094P.  
XX 10-JUN-1999; 99US-0138540P.  
XX 14-JUN-1999; 99US-0138847P.  
XX 16-JUN-1999; 99US-0139119P.  
XX 16-JUN-1999; 99US-0139452P.  
XX 17-JUN-1999; 99US-0139453P.  
XX 18-JUN-1999; 99US-0139454P.  
XX 18-JUN-1999; 99US-0139455P.



PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 20-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147030P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.

PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159283P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

## Query Match

100.0%; Score 25; DB 3; Length 61;

Best Local Similarity 50.0%; Pred. NO. 9.6e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8

Db 44 LEFIEDE 51

## RESULT 35

ABG00521

ID ABG00521 standard; protein; 64 AA.

XX

AC ABG00521;  
XX DT 13-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #512.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PR 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-ESDB; AAS64708.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX PS Claim 20; SEQ ID NO 30880; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 64 AA;  
  
Query March 100.0%; Score 25; DB 4; Length 64;  
Best Local Similarity 50.0%; Pred. No.1e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 XEFTXXDX 8  
Db :|||:|:  
44 QEFINLDG 51  
  
RESULT 36  
AAG24528  
ID AAG24528 standard; protein; 66 AA.  
XX  
AC AAG24528;  
XX

DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 28237.  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 28237.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EPI033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.

```
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 28-JUN-1999; 99US-0140695P.
PR 29-JUN-1999; 99US-0140823P.
PR 30-JUN-1999; 99US-0140951P.
PR 01-JUL-1999; 99US-0141387P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143342P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
```

```
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160800P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161320P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
```

Query Match 100.0%; Score 25; DB 3; Length 66;  
Best Local Similarity 50.0%; Pred. No. 1e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
Db :|||::|:  
40 IEFICLDL 47

## RESULT 37

AAG10475  
ID AAG10475 standard; protein; 66 AA.  
XX  
AC AAG10475;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8814.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

xx Arabidopsis thaliana.  
OS  
XX  
PN EP1033405-A2.  
XX  
PD  
PF  
PF  
XX  
XX  
PR 25-FEB-1999; 99US-011825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 29-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131443P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149923P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 31-AUG-1999; 99US-0151303P.  
PR 01-SEP-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.

PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 08-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159233P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161320P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 66;  
Best Local Similarity 50.0%; Pred. No. 1e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 XEFIXDX 8  
Db 40 IEFICLDL 47

RESULT 38  
ABP80025  
ID ABP80025 standard; protein; 67 AA.  
XX AC  
XX ABP80025;  
XX DT  
XX 07-MAR-2003 (first entry)

DE N. gonorrhoeae amino acid sequence SEQ ID 6580.  
XX Antibacterial; infection; vaccine; gene therapy.  
KW Neisseria gonorrhoeae.  
XX OS  
XX WC200279243-A2.  
XX PN  
XX 10-OCT-2002.  
XX PD  
XX 12-FEB-2002; 2002WO-IB002069.  
XX PF  
XX

PR 12-FEB-2001; 2001GB-00003424.  
XX (CHIR-) CHIRON SPA.  
XX PI

Fontana MR, Pizza M, Masighani V, Monaci E;

WPI; 2003-058415/05.  
DR N-PSDB; ABZ40995.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
PT medicament for treating or preventing N. gonorrhoeae infection.  
XX PS

Disclosure; Page 666; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae.  
CC Also disclosed are the nucleic acid molecules encoding the proteins and  
CC antibodies that specifically bind to the proteins. The composition  
CC comprising the protein, nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
CC infection, this may be in the form of a vaccine or gene therapy.  
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
XX molecules of the invention

SQ Sequence 67 AA;

Query Match 100.0%; Score 25; DB 6; Length 67;  
Best Local Similarity 50.0%; Pred. No. 1.1e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 XEFIXDX 8  
Db 16 HEFIYDPS 23

RESULT 39  
AAU16600  
ID AAU16600 standard; protein; 68 AA.  
XX AC

AAU16600;

XX 07-NOV-2001 (first entry)

XX Human novel secreted protein, Seq ID 1553.

XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;  
KW antibacterial; virucide; fungicide; ophthalmological; vulnery;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
KW preservative; antiproliferative.

XX Homo sapiens.

XX WO200155322-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001341.

XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0232402P.  
 PR 14-SEP-2000; 2000US-0232403P.  
 PR 14-SEP-2000; 2000US-0232404P.  
 PR 14-SEP-2000; 2000US-0232405P.  
 PR 14-SEP-2000; 2000US-0232406P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234224P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 25-SEP-2000; 2000US-0234999P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.

PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249246P.  
 PR 17-NOV-2000; 2000US-0249256P.  
 PR 17-NOV-2000; 2000US-0249257P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251858P.  
 PR 08-DEC-2000; 2000US-0251859P.  
 PR 08-DEC-2000; 2000US-0251899P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-488783/53.  
 XX N-PSDB; AAS26587.  
 DR  
 XX  
 PT  
 PT  
 XX  
 RS  
 XX  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC

New nucleic acid molecules encoding 461 human secreted proteins for diagnosing, preventing, treating or ameliorating medical conditions and used as food additives or preservatives.  
 Claim 11; SEQ ID NO 1553; 980pp; English.  
 The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic

CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays  
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodenesis,  
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
 CC and many other disorders listed in the specification. The polypeptides  
 CC can also be used to aid wound healing and epithelial cell proliferation,  
 CC to prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemoraxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed

Query Match 100.0%; Score 25; DB 4; Length 68;  
 Best Local Similarity 50.0%; Pred. No. 1.1e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 Db 11 EEFIGDR 18

RESULT 40  
 ABP64248  
 ID ABP64248 standard; protein; 68 AA.

AC ABP64248;

DT 04-NOV-2002 (first entry)

DE Human ORF618.

XX Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;  
 KW Antiinflammatory; Gene therapy; human; ORFX; atherogenic; platelet;  
 KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;  
 KW cancer; cardiovascular disease; allergy; autoimmune disease;  
 KW wound healing; blood coagulation disorder; inflammatory disorder.

XX Homo sapiens.

PN US2002082206-A1.

XX 27-JUN-2002.

PF 30-MAY-2001; 2001US-00867550.

PR 30-MAY-2000; 2000US-0208427P.

XX (LEAC/) LEACH M D.

PA (MEHR/) MEHRABAN F.

PA (CONLEY/) CONLEY P B.

PA (TOPP/) TOPPER J N.

PA (LAWD/) LAW D.

XX Leach MD, Mehraban F, Conley PB, Topper JN, Law D;

XX WPI; 2002-626554/67.

DR N-PSDB; ABQ98811.

XX New polypeptide designated ORFX are present in human atherogenic cells  
 PT and are useful to prevent and treat ORFX-associated disorders including  
 PT cancer, allergy, wound healing or autoimmune, cardiovascular or  
 PT inflammatory disease.

XX Claim 10; SEQ ID NO 1236; 78pp; English.

XX The present invention relates to novel human ORFX polypeptides and their  
 CC coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences

CC were discovered in human atherogenic cells, in particular in platelets  
 CC and human umbilical vein endothelial cells (HUVEC) and are expressed in  
 CC many other tissues as well. Atherogenic cells are cells which have the  
 CC potential to develop atherosclerotic plaques. The ORFX polypeptides and  
 CC nucleic acids are useful for treating or preventing a pathological  
 CC condition associated with an ORFX-associated disorder, e.g. cancer,  
 CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood  
 CC coagulation disorders or inflammatory disorders. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from the USPTO web site at  
 CC seqdata.uspto.gov/sequence.html?DocID=20020082206

XX SQ Sequence 68 AA;

Query Match 100.0%; Score 25; DB 5; Length 68;  
 Best Local Similarity 50.0%; Pred. No. 1.1e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 Db 21 CEFIKDK 28

RESULT 41  
 ABU55669  
 ID ABU55669 standard; protein; 68 AA.

XX AC ABU55669;

DT 18-MAR-2003 (first entry)

DE Human novel polypeptide #756.

XX Human; neural disorder; immune system disorder; renal disorder;  
 KW muscular disorder; respiratory disease; reproductive disorder;  
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
 KW blood related disorder; cancer; immunosuppressive; antinflammatory;  
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
 KW haemostatic; antiarteriosclerotic.

XX Homo sapiens.

PN US2002132753-A1.

XX 19-SEP-2002.

XX 17-JAN-2001; 2001US-00764864.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 28-JUN-2000; 2000US-0214886P.

PR 07-JUL-2000; 2000US-0215647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 22-AUG-2000; 2000US-0226868P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-02344223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 XX (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 DR WPI; 2003-147444/14.  
 DR N-PSDB; ABX73928.  
 XX  
 PT New polypeptides and nucleic acids, useful in gene therapy for treating,  
 PT inhibiting or preventing e.g. neural, immune system, muscular,  
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
 PT renal disorders.  
 XX  
 PS Claim 11; SEQ ID NO 1553; 402pp; English.  
 XX  
 CC The invention relates to human novel polypeptides and their associated  
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
 CC therapy for treating, inhibiting or preventing neural disorders, immune  
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left  
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and  
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood  
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
 CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and  
 CC ABU55748 represent human novel polypeptides of the invention  
 XX  
 SQ Sequence 68 AA;  
 Query Match 100.0%; Score 25; DB 6; Length 68;  
 Best Local Similarity 50.0%; Pred. No. 1.1e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEFIXDX 8  
 :|||:|:  
 Db 11 EEFIGDGR 18  
 RESULT 42  
 AAG01116  
 ID AAG01116 standard; protein; 69 AA.

XX AAG01116;  
 AC 06-OCT-2000 (first entry)  
 DT Human secreted protein, SEQ ID NO: 5197.  
 DE Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.  
 KW Homo sapiens.  
 OS EP1033401-A2.  
 PN 06-SEP-2000.  
 PD 21-FEB-2000; 2000EP-00200610.  
 XX 26-FEB-1999; 99US-0122487P.  
 XX (GEST ) GENSET.  
 PA Dumas Milne Edwards J, Duclert A, Giordano J;  
 PI WPI; 2000-500381/45.  
 DR N-PSDB; AAC01122.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 XX  
 PS Claim 13; SEQ ID NO 5197; 71pp + Sequence Listing; English.  
 XX  
 CC The present sequence is a polypeptide encoded by one of a large number of  
 CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
 CC prepared from total human RNAs or polyA+ RNAs derived from 30 different  
 CC tissues. EST sequences usually correspond mainly to the 3' untranslated  
 CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
 CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
 CC sequences derived from the 5' ends of mRNAs and even in those cases where  
 CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
 CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
 CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
 CC are also used in diagnostic, forensic, gene therapy and chromosome  
 CC mapping procedures. They are used to obtain upstream regulatory sequences  
 CC and to design expression and secretion vectors  
 XX  
 SQ Sequence 69 AA;  
 Query Match 100.0%; Score 25; DB 3; Length 69;  
 Best Local Similarity 50.0%; Pred. No. 1.1e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEFIXDX 8  
 :|||:|:  
 Db 32 VEFIRHDR 39  
 RESULT 43  
 ABP77688  
 ID ABP77688 standard; protein; 69 AA.  
 XX  
 AC ABP77688;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE N. gonorrhoeae amino acid sequence SEQ ID 1906.  
 XX Antibacterial; infection; vaccine; gene therapy.  
 KW  
 XX Neisseria gonorrhoeae.  
 OS  
 XX WO200279243-A2.  
 FN





CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage. N.B. The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 73 AA;

Query Match 100.0%; Score 25; DB 5; Length 73;  
 Best Local Similarity 50.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||:|:  
 Db 2 LEFIPSDQ 9

RESULT 46  
 AAW61649  
 ID AAW61649 standard; peptide; 74 AA.  
 XX  
 AC AAW61649;  
 XX  
 DT 27-OCT-1998 (first entry)  
 XX  
 DE Non-ocular disease marker 3.  
 XX  
 KW Human; non-ocular disease; tear; cancer; breast; prostate.  
 XX  
 OS Homo sapiens.  
 XX  
 FN W09835229-A1.  
 XX  
 PD 13-AUG-1998.  
 XX  
 PF 06-FEB-1998; 98WO-AU0000071.  
 XX  
 PR 07-FEB-1997; 97AU-00005009.  
 XX  
 PA (MACQ-) MACQUARIE RES LTD.  
 PA (UNIX) UNISEARCH LTD.  
 XX  
 PI Morris C, Wallcox M, Bolis S, Walsh B, Herbert B, Molloy M;  
 PI Gooley AA, Williams KL;  
 XX  
 DR WPI; 1998-447373/38.

PT Screening for non-ocular disease - by analysing tears for marker  
 PT proteins, particularly indicative of cancer and genetic disease, also new  
 PT proteins and nucleic acid encoding them.

PS Claim 6; Page 9; 14pp; English.

XX The markers AAW61647-W61649 are used for screening for, or detecting, non  
 CC ocular disease by analysing tears. Biochemicals, specifically proteins,  
 CC are isolated from tears, particularly by chromatography or  
 CC electrophoresis, especially two-dimensional polyacrylamide gel  
 CC electrophoresis (2D-PAGE), then detected, e.g. with labelled specific  
 CC reagents, in (radio)immunoassay. The method is used to detect cancer,

CC particularly of breast or prostate, or a genetic disease, in humans or  
 CC animals  
 XX Sequence 74 AA;  
 SQ

Query Match 100.0%; Score 25; DB 2; Length 74;  
 Best Local Similarity 50.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||:|:  
 Db 29 QEFIDSDA 36

RESULT 47

ABP38159  
 ID ABP38159 standard; protein; 80 AA.

XX  
 AC ABP38159;  
 XX  
 DT 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3004.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-00134001.

XX 14-AUG-1997; 97US-0055779P.

XX 08-NOV-1997; 97US-0064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

XX N-PSDB; ABN90704.

PT Novel isolated nucleic acid encoding a Staphylococcus epidermis  
 PT polypeptide, useful for diagnosing and treating bacterial infections.

XX Disclosure; SEQ ID NO 3004; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life cycle  
 CC or inhibit S. epidermidis infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site

XX Sequence 80 AA;

Query Match 100.0%; Score 25; DB 5; Length 80;  
 Best Local Similarity 50.0%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||:|:  
 Db 54 EEFINDDS 61

RESULT 48

```

ADA36717
ID ADA36717 standard; protein; 80 AA.
XX
AC ADA36717;
XX
DT 20-NOV-2003 (first entry)
XX
DE Acinetobacter baumannii protein #3878.
XX
KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.
DR N-PSDB; ADA32591.
XX
PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 8004; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
SQ Sequence 80 AA;
Query Match 100.0%; Score 25; DB 6; Length 80;
Best Local Similarity 50.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db 30 DEFIKDX 37
RESULT 49
AAW62670
ID AAW62670 standard; protein; 82 AA.
XX
AC AAW62670;
XX
DT 09-NOV-1998 (first entry)
XX
DE Streptococcus pneumoniae polypeptide.
XX
KW Polypeptide; ORF; open reading frame; infection; bacterial;
KW streptococcal; bacteremia; diagnosis; prophylaxis.
XX
OS Streptococcus pneumoniae.
XX
PN WO9823631-A1.
ADA36717
ID ADA36717 standard; protein; 80 AA.
XX
AC ADA36717;
XX
DT 20-NOV-2003 (first entry)
XX
DE Acinetobacter baumannii protein #3878.
XX
KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.
DR N-PSDB; ADA32591.
XX
PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 8004; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
SQ Sequence 80 AA;
Query Match 100.0%; Score 25; DB 6; Length 80;
Best Local Similarity 50.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db 30 DEFIKDX 37
RESULT 49
AAW62670
ID AAW62670 standard; protein; 82 AA.
XX
AC AAW62670;
XX
DT 09-NOV-1998 (first entry)
XX
DE Streptococcus pneumoniae polypeptide.
XX
KW Polypeptide; ORF; open reading frame; infection; bacterial;
KW streptococcal; bacteremia; diagnosis; prophylaxis.
XX
OS Streptococcus pneumoniae.
XX
PN WO9823631-A1.
ADA36717
ID ADA36717 standard; protein; 80 AA.
XX
AC ADA36717;
XX
DT 20-NOV-2003 (first entry)
XX
DE Acinetobacter baumannii protein #3878.
XX
KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US6562958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.
DR N-PSDB; ADA32591.
XX
PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
PS Example; SEQ ID NO 8004; 328pp; English.
XX
CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
SQ Sequence 80 AA;
Query Match 100.0%; Score 25; DB 2; Length 82;
Best Local Similarity 50.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db 58 NEFIALDD 65
RESULT 50
ABP08611
ID ABP08611 standard; protein; 82 AA.
XX
AC ABP08611;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:17204.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;

```

XX WPI: 2002-106308/14.  
 DR N-PSDB; ABN24363.  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders.  
 XX  
 PS Disclosure; SEQ ID NO 17204; 1037pp; English.  
 XX  
 CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage. N.B. The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 82 AA;  
 Query Match 100.0%; Score 25; DB 5; Length 82;  
 Best Local Similarity 50.0%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEFIXDX 8  
 Db :|||:|:  
 62 REFICIDT 69  
 RESULT 51  
 AAB83324  
 ID AAB83324 standard; protein; 83 AA.  
 XX  
 AC AAB83324;  
 XX  
 DT 09-OCT-2001 (first entry)  
 XX  
 XX Rice isoleucyl-tRNA synthetase protein sequence.  
 DE  
 KW Aminoacyl-tRNA synthetase; isoleucyl-tRNA synthetase; transgenic plant;  
 KW lysyl-tRNA synthetase; phenylalanyl-tRNA synthetase; gene expression;  
 KW prolyl-tRNA synthetase; herbicide identification; genetic marker;  
 KW plant breeding.  
 XX  
 OS Oryza sativa.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 76 /note= "encoded by NGG"  
 FT  
 XX US6271441-B1.  
 PN  
 XX 07-AUG-2001.  
 PD  
 XX 20-JUL-1999; 99US-00357251.  
 PF  
 XX

PR 21-JUL-1998; 98US-0093530P.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Falco SC, Famodu LO, Orozco EM, Schwaber JS;  
 XX  
 DR WPI: 2001-482449/52.  
 DR N-PSDB; AAF87079.  
 XX  
 PT Isolated polynucleotide encoding plant isoleucyl-tRNA synthase, useful  
 PT for creating transgenic plants and as targets to facilitate design and/or  
 PT identification of inhibitors that may be useful as herbicides.  
 XX  
 PS Claim 1; Col 25-28; 50pp; English.  
 XX  
 CC This sequence is a plant aminoacyl-tRNA synthetase of the invention. The  
 CC aminoacyl-tRNA synthetases are selected from isoleucyl-tRNA synthetase,  
 CC lysyl-tRNA synthetase, phenylalanyl-tRNA synthetase, and prolyl-tRNA  
 CC synthetase. The nucleic acid fragments may be used to create transgenic  
 CC plants in which the polypeptides are present at higher or lower levels  
 CC than normal or in cell types or developmental stages in which they are  
 CC not normally found. This would have the effect of altering the level of  
 CC aminoacyl-tRNA synthetase activity and gene expression in those cells.  
 CC The polypeptides can be used as a targets to facilitate design and/or  
 CC identification of inhibitors of those enzymes that may be useful as  
 CC herbicides. All or a substantial portion of the nucleic acid fragments  
 CC may also be used as probes for genetically and physically mapping the  
 CC genes that they are a part of, and as markers for traits linked to those  
 CC genes. Such information may be useful in plant breeding in order to  
 CC develop lines with desired phenotypes  
 XX  
 SQ Sequence 83 AA;  
 Query Match 100.0%; Score 25; DB 4; Length 83;  
 Best Local Similarity 50.0%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEFIXDX 8  
 Db :|||:|:  
 31 EEFIFYDG 38  
 RESULT 52  
 ABU08739  
 ID ABU08739 standard; protein; 83 AA.  
 XX  
 AC ABU08739;  
 XX  
 DT 25-JUN-2003 (first entry)  
 XX  
 XX Rice isoleucyl-tRNA synthetase clone rls2.pk0006.c10.  
 DE  
 KW Rice; isoleucyl-tRNA synthetase; lysyl-tRNA synthetase;  
 KW phenylalanyl-tRNA synthetase; prolyl-tRNA synthetase;  
 KW aminoacyl-tRNA synthetase; herbicide; enzyme.  
 XX  
 OS Oryza sativa.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 76 /label= Unknown  
 FT  
 XX US2003018985-A1.  
 XX  
 PD 23-JAN-2003.  
 XX  
 PF 24-MAY-2001; 2001US-00864464.  
 XX  
 XX 21-JUL-1998; 98US-0093530P.  
 PR 20-JUL-1999; 99US-00357251.  
 XX  
 XX (FALC/) FALCO S C.  
 PA (FAWO/) FAMODU O O.  
 PA

```

PA (SIMM/) SIMMONS C R.
XX PI Falco SC, Farnodu OO, Simmons CR;
XX WPI; 2003-401721/38.
XX DR N-PSDB; ABX95455.
XX PT New nucleic acid fragments encoding aminoacyl-tRNA synthetases, useful in
XX developing new herbicides that target aminoacyl-tRNA synthetases and
XX engineer aminoacyl-tRNA synthetases that are resistant to the herbicides.
XX Claim 6; Page 15; 54pp; English.
XX CC The invention describes an isolated nucleic acid fragment encoding an
XX isoleucyl-tRNA synthetase, a lysyl-tRNA synthetase, a phenylalanyl-tRNA
XX synthetase, or a prolyl-tRNA synthetase. The nucleic acid fragments
XX encoding aminoacyl-tRNA synthetases are useful in developing new
XX herbicides that target aminoacyl-tRNA synthetases and engineer aminoacyl-
XX tRNA synthetases that are resistant to the herbicides. The nucleic acid
XX fragments are also useful in facilitating studies to better understand
XX protein synthesis in plants, providing genetic tools for the manipulation
XX of gene expression, or providing possible target for herbicides. The
XX polypeptides are useful as targets to facilitate design and/or
XX identification of inhibitors of the aminoacyl-tRNA synthetases that may
XX be used as herbicides. This is the amino acid sequence of rice isoleucyl-
XX tRNA synthetase clone rls2.pk0006.c10
XX SQ Sequence 83 AA;
Query Match 100.0%; Score 25; DB 6; Length 83;
Best Local Similarity 50.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db 31 EEFIFYDG 38
RESULT 53
AAG18654
ID AAG18654 standard; protein; 85 AA.
XX AC AAG18654;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 20148.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
XX PR 05-MAR-1999; 99US-0123180P.
XX PR 03-MAR-1999; 99US-0123548P.
XX PR 23-MAR-1999; 99US-0125788P.
XX PR 25-MAR-1999; 99US-0126264P.
XX PR 29-MAR-1999; 99US-0126785P.
XX PR 01-APR-1999; 99US-0127462P.
XX PR 06-APR-1999; 99US-0128234P.
XX PR 08-APR-1999; 99US-0128714P.
XX PR 16-APR-1999; 99US-0129845P.
XX PR 19-APR-1999; 99US-0130077P.
XX PR 21-APR-1999; 99US-0130449P.
XX PR 23-APR-1999; 99US-0130510P.
XX PR 23-APR-1999; 99US-0130891P.
XX PR 28-APR-1999; 99US-0131449P.
XX PR 30-APR-1999; 99US-0132048P.
XX PR 30-APR-1999; 99US-0132407P.
XX PR 04-MAY-1999; 99US-0132484P.
XX PR 05-MAY-1999; 99US-0132485P.
XX PR 06-MAY-1999; 99US-0132486P.
XX PR 06-MAY-1999; 99US-0132487P.
XX PR 07-MAY-1999; 99US-0132863P.
XX PR 11-MAY-1999; 99US-0134256P.
XX PR 14-MAY-1999; 99US-0134218P.
XX PR 14-MAY-1999; 99US-0134219P.
XX PR 14-MAY-1999; 99US-0134221P.
XX PR 14-MAY-1999; 99US-0134370P.
XX PR 18-MAY-1999; 99US-0134768P.
XX PR 19-MAY-1999; 99US-0134941P.
XX PR 20-MAY-1999; 99US-0135124P.
XX PR 21-MAY-1999; 99US-0135353P.
XX PR 24-MAY-1999; 99US-0135629P.
XX PR 25-MAY-1999; 99US-0136021P.
XX PR 27-MAY-1999; 99US-0136392P.
XX PR 28-MAY-1999; 99US-0136782P.
XX PR 01-JUN-1999; 99US-0137222P.
XX PR 03-JUN-1999; 99US-0137528P.
XX PR 04-JUN-1999; 99US-0137502P.
XX PR 07-JUN-1999; 99US-0137724P.
XX PR 08-JUN-1999; 99US-0138094P.
XX PR 10-JUN-1999; 99US-0138540P.
XX PR 10-JUN-1999; 99US-0138847P.
XX PR 14-JUN-1999; 99US-0139119P.
XX PR 16-JUN-1999; 99US-0139452P.
XX PR 16-JUN-1999; 99US-0139453P.
XX PR 17-JUN-1999; 99US-0139492P.
XX PR 18-JUN-1999; 99US-0139454P.
XX PR 18-JUN-1999; 99US-0139455P.
XX PR 18-JUN-1999; 99US-0139456P.
XX PR 18-JUN-1999; 99US-0139457P.
XX PR 18-JUN-1999; 99US-0139458P.
XX PR 18-JUN-1999; 99US-0139459P.
XX PR 18-JUN-1999; 99US-0139460P.
XX PR 18-JUN-1999; 99US-0139461P.
XX PR 18-JUN-1999; 99US-0139462P.
XX PR 18-JUN-1999; 99US-0139463P.
XX PR 18-JUN-1999; 99US-0139750P.
XX PR 18-JUN-1999; 99US-0139763P.
XX PR 21-JUN-1999; 99US-0139817P.
XX PR 22-JUN-1999; 99US-0139899P.
XX PR 23-JUN-1999; 99US-0140353P.
XX PR 23-JUN-1999; 99US-0140354P.
XX PR 24-JUN-1999; 99US-0140695P.
XX PR 28-JUN-1999; 99US-0140823P.
XX PR 29-JUN-1999; 99US-0140991P.
XX PR 30-JUN-1999; 99US-0141287P.
XX PR 01-JUL-1999; 99US-0141842P.
XX PR 01-JUL-1999; 99US-0142154P.
XX PR 02-JUL-1999; 99US-0142055P.
XX PR 06-JUL-1999; 99US-0142390P.
XX PR 08-JUL-1999; 99US-0142803P.
XX PR 09-JUL-1999; 99US-0142920P.
XX PR 12-JUL-1999; 99US-0142977P.
XX PR 13-JUL-1999; 99US-0143542P.
XX PR 14-JUL-1999; 99US-0143624P.
XX PR 15-JUL-1999; 99US-0144005P.
XX PR 16-JUL-1999; 99US-0144085P.
XX PR 16-JUL-1999; 99US-0144086P.
XX PR 19-JUL-1999; 99US-0144325P.
XX PR 19-JUL-1999; 99US-0144331P.
XX PR 19-JUL-1999; 99US-0144332P.
XX PR 19-JUL-1999; 99US-0144333P.
XX PR 19-JUL-1999; 99US-0144334P.
XX PR 19-JUL-1999; 99US-0144335P.
XX PR 20-JUL-1999; 99US-0144352P.
XX PR 20-JUL-1999; 99US-0144632P.

```

PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 21-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 08-OCT-1999; 99US-0157753P.  
PR 08-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159285P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.

PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159594P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 29-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 85;  
Best Local Similarity 50.0%; Pred. No. 1.4e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFLXDX 8  
Db 13 PEFIFEDG 20  
:|||||:

## RESULT 54

ABB90813  
ID ABB90813 standard; protein; 89 AA.  
XX AC ABB90813;  
XX 31-MAY-2002 (first entry)  
XX Herbicidally active polypeptide SEQ ID NO 24.  
DE Herbicidally active polypeptide SEQ ID NO 24.  
XX Herbicidally active polypeptide SEQ ID NO 24.  
XX Arabidopsis thaliana.  
XX WO200210210-A2.  
XX 07-FEB-2002.  
XX 28-AUG-2001; 2001WO-EP009892.  
XX 28-AUG-2001; 2001WO-EP009892.  
XX (FARB ) BAYER AG.  
XX Tietjen K, Weidler M;  
XX WPI; 2002-269010/31.  
XX Identifying plant target proteins for herbicidally active compounds,  
PT comprising aligning and comparing nucleic acid or amino acid sequences,  
PT from plant with nucleic acid or amino acid sequences from non-plant  
PT organisms.  
XX Claim 5; SEQ ID NO 24; 261pp + Sequence Listing; English.  
XX The invention relates to identifying target proteins (ABB90790-ABB94016)  
CC for herbicidally active compounds, comprising aligning and comparing  
CC nucleic acid or amino acid sequences from plant with nucleic acid or  
CC amino acid sequences from non-plant organisms using suitable search  
CC parameters, where plant sequences having an E-value greater by a factor  
CC of 3 than the E-value of most similar non-plant sequences are selected.  
CC The polypeptides or nucleic acids encoding them are useful for

CC identifying modulators. The identified modulators are useful as  
 CC herbicides

SQ Sequence 89 AA;

Query Match 100.0%; Score 25; DB 5; Length 89;  
 Best Local Similarity 50.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 XEFIXDX 8

:|||::|:

Db 52 TERIRSDI 59

RESULT 55

AAP81887

ID AAP81887 standard; protein; 90 AA.

AC AAP81887;

XX 25-MAR-2003 (revised)

DT 31-DEC-1990 (first entry)

XX Sequence of human prolaxin B chain.

XX Hormone; relaxin analogue; antagonist.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Region 11..27

FT Peptide /label= B CHAIN

FT 28..28

FT /label= C PEPTIDE

PN EP287820-A.

XX 26-OCT-1988.

XX 11-AUG-1983; 83EP-00104503.

XX 12-AUG-1982; 82AU-00005352.

XX 23-APR-1987; 88EP-00104503.

XX (FLOR-) FLOREY INST EXP PHY.

XX Hudson PJ, Shine J, Niall HD, Tregear GW;

XX WPI; 1988-300910/43.

XX N-PSDB; AAN81773.

XX Human relaxin analogues - has shortened and/or modified forms of natural  
 PT B and/or A chains modified by addn. of protective gp. to free amino gp.

PS Disclosure; Fig 2B; 22pp; English.

XX Human relaxin may be prep'd. by combining the A and B chains of relaxin in  
 CC full-length, shortened or modified forms. Pref'd. analogues consist of any  
 CC one of the A-chains A(1-24), A(2-24), A(3-24) in combination with any of  
 CC the B-chains B(1-23) to B(1-32). (Updated on 25-MAR-2003 to correct PR  
 CC field.)

XX Sequence 90 AA;

Query Match 100.0%; Score 25; DB 1; Length 90;

Best Local Similarity 50.0%; Pred. No. 1.4e+03;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 XEFIXDX 8

:|||::|:

Db 29 QEFINKDT 36

RESULT 56

ABP07728

ID ABP07728 standard; protein; 94 AA.

XX ABP07728;

XX 25-JUN-2002 (first entry)

XX Human ORFX protein sequence SEQ ID NO:15438.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.

XX Homo sapiens.

XX WO200192523-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US010836.

XX 30-MAY-2000; 2000US-0206132P.

XX 29-AUG-2000; 2000US-0228716P.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach MD;

XX WPI; 2002-106308/14.

XX N-PSDB; ABN23480.

XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders.

XX Disclosure; SEQ ID NO 15438; 1037pp; English.

XX The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage. N.B. The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 94 AA;

Query Match 100.0%; Score 25; DB 5; Length 94;

Best Local Similarity 50.0%; Pred. No. 1.5e+03;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 XEFIXDX 8
Db :|||:|:
47 QEFIDSDA 54

RESULT 57
AAW35804
ID AAW35804 standard; protein; 95 AA.
XX
AC AAW35804;
XX
DT 27-MAR-1998 (first entry)
XX
DE Human endometrial specific steroid-binding factor III.
XX
KW Endometrial specific steroid-binding factor III; ESF III; human;
KW Clara cell secretory protein; endometrium; phospholipase A2 inhibitor;
KW polychlorinated biphenyl; antaggregant; inflammation; asthma; rhinitis;
KW cystic fibrosis; airway disease; neoplasia; atopy; therapy; diagnosis.
XX
OS Homo sapiens.
XX
FN WO9734997-A1.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= Sig_peptide
FT Protein 22..95
FT /label= Mat_protein
XX
PN WO9734997-A1.
XX
PD 25-SEP-1997.
XX
PF 21-MAR-1996; 96WO-US003857.
XX
PR 21-MAR-1996; 96WO-US003857.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Yu G, Gentz RL;
XX
WPI; 1997-480206/44.
DR N-PSDB; AAT94832.
XX
Human endometrial specific steroid-binding factor I, II and III - used to
treat inflammation, asthma, rhinitis, cystic fibrosis, airway disease,
neoplasia, atopy etc.
XX
Claim 19; Page 65; 92pp; English.
XX
This sequence comprises human endometrial specific steroid binding factor
III (ESF III), a protein that inhibits phospholipase A2 activity, binds
to polychlorinated biphenyl compounds, reduces foreign protein
antigenicity, inhibits monocyte and neutrophil chemotaxis and
phagocytosis, inhibits platelet aggregation, regulates eicosanoid levels
in the human uterus and controls the growth of endometrial cells. The
amino acid sequence was deduced from a cDNA clone (see AAT94832) derived
from a human endometrial tumour. ESF I (see AAW35802) and ESF II (see
AAW35803) are also claimed. Human ESF III has about 36% identity with rat
prostatic steroid-binding protein. Recombinant ESF I, II and III can be
expressed in host cells for use in claimed methods (a) for treating a
patient in need of ESF I, II or III (including expression of the
polypeptide in vivo) and (b) for identifying compounds which bind to and
inhibit activation of the ESF polypeptide. hESF I, II and III may be used
to treat inflammation, asthma, rhinitis, cystic fibrosis, airway disease,
neoplasia and atopy
XX
SQ Sequence 95 AA;
Query Match 100.0%; Score 25; DB 2; Length 95;
Best Local Similarity 50.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 XEFIXDX 8
Db :|||:|:
47 QEFIDSDA 54

RESULT 58
AAW02590
ID AAW02590 standard; protein; 95 AA.
XX
AC AAW02590;
XX
DT 26-JUL-1999 (first entry)
XX
DE A human mamoglobin homologue (HMH).
XX
KW Human mamoglobin homologue; HMH; antagonist; neoplastic disorder;
KW adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;
KW teratocarcinoma; endometriosis.
XX
OS Homo sapiens.
XX
FN WO9919487-A1.
XX
PD 22-APR-1999.
XX
PF 14-OCT-1998; 98WO-US021729.
XX
PR 16-OCT-1997; 97US-00951750.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Hillman JL, Shah P, Murry LE;
XX
WPI; 1999-302531/25.
DR N-PSDB; AAX36138.
XX
New human mamoglobin homolog (HMH), useful for diagnosing, treating or
preventing disorders associated with expression of HMH.
XX
Claim 1; Fig 1A-B; 63pp; English.
XX
The present sequence represents a human mamoglobin homologue (HMH).
Antagonists of the HMH polypeptide can be used to treat neoplastic
disorders including adenocarcinoma, leukemia, lymphoma, melanoma,
myeloma, sarcoma and teratocarcinoma. A vector expressing the complement
of the polynucleotide encoding HMH may be administered to a subject to
treat or prevent neoplastic disorders or endometriosis. Antibodies which
bind HMH may also be used in the diagnosis of conditions or diseases
characterized by expression of HMH, or in assays to monitor patients
being treated with HMH, agonists, antagonists or inhibitors.
XX
Polynucleotides encoding HMH may also be used diagnostically to detect
and quantitate gene expression in biopsied tissues. With respect to
cancer a relatively high amount of transcript may indicate a
predisposition for the development of disease. The nucleic acid sequences
which encode HMH may also be used to generate hybridization probes useful
for mapping the naturally occurring genomic sequence. HMH, and its
fragments/variants can be used for screening libraries of compounds in
drug screening techniques
XX
SQ Sequence 95 AA;
Query Match 100.0%; Score 25; DB 2; Length 95;
Best Local Similarity 50.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 XEFIXDX 8
Db :|||:|:
47 QEFIDSDA 54

RESULT 59
AAB03769
ID AAB03769 standard; protein; 95 AA.
```



```

XX AC AAB03769;
XX AC
XX DT 06-OCT-2000 (first entry)
XX DE
XX DE Human endometrial specific steroid-binding factor III protein sequence.
XX KW Endometrial specific steroid-binding factor; human; hESF; inflammation;
XX KW asthma; rhinitis; cystic fibrosis; air way disease; neoplasia; atopy;
XX KW eicosanoid level regulator; chemotaxis inhibitor; endometrial cancer; ss.
XX OS Homo sapiens.
XX OS
XX PN US0666724-A.
XX PD
XX PD 23-MAY-2000.
XX PF 21-MAR-1997; 97US-00821451.
XX PR 21-MAR-1996; 96US-0014724P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Yu G, Gentz R, Ni J;
XX DR WPI; 2000-375600/32.
XX DR N-PSDB; AAA59730.
XX PT Novel gene encoding human endometrial specific steroid-binding factor I,
XX PT II and III which is useful for treating asthma, rhinitis, cystic
XX PT fibrosis, airway disease and neoplasia.
XX PS Claim 1; Fig 3; 36pp; English.
XX CC This invention relates to nucleic acid molecules encoding portions of the
XX CC human endometrial specific steroid-binding factors I, II, and III. Also
XX CC included in the invention are hESF I, II, and III polypeptide sequences.
XX CC The nucleotide sequence exhibit antiasthmatic, antiinflammatory,
XX CC anti-allergic, and cytostatic properties. The polynucleotides are used in
XX CC gene therapy to express hESF I, II and III polypeptides in vivo to treat
XX CC and/or prevent inflammation, asthma, rhinitis, cystic fibrosis, air way
XX CC disease, neoplasia and atopy. The polynucleotides are also used to
XX CC inhibit phospholipase A2 activity, bind polychlorinated biphenyls, reduce
XX CC foreign protein antigenicity, inhibit monocyte and neutrophil chemotaxis
XX CC and phagocytosis, inhibit platelet aggregation, regulate eicosanoid
XX CC levels in the human uterus and control the growth of endometrial cells.
XX CC The polynucleotides are also useful for detecting complementary
XX CC polynucleotides as a diagnostic reagent. The hESF I, II and III
XX CC polynucleotides are used to detect complementary polynucleotides such as
XX CC a diagnostic reagent. Detection of a mutated form of hESF I, II and III
XX CC associated with a dysfunction will provide a diagnostic tool that can
XX CC define diagnosis of a disease or susceptibility to a disease which
XX CC results from under-expression, over-expression or altered expression of
XX CC hESF I, II and III e.g. a susceptibility to inherited asthma and
XX CC endometrial cancer. They are also useful for chromosome identification.
XX CC The present invention
XX CC represents a hESF III protein sequence identified in
XX CC the invention
XX CC
XX CC Sequence 95 AA;
XX CC
XX CC Query Match 100.0%; Score 25; DB 3; Length 95;
XX CC Best Local Similarity 50.0%; Pred. No. 1.5e+03;
XX CC Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
XX CC
XX CC 1 XEFIXDX 8
XX CC :|||:|:|:
XX CC 47 QEFIDSDA 54
XX CC
XX CC RESULT 60
XX CC AAY65394
XX CC ID AAY65394 standard; protein; 95 AA.
XX CC

```

```

AC AAY65394;
XX 01-FEB-2000 (first entry)
XX Human 5' EST related polypeptide SEQ ID NO:1555.
XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
XX gene therapy; chromosome mapping; upstream regulatory sequence; forensic;
XX location; development; protein synthesis; stability; regulation;
XX identification.
XX OS Homo sapiens.
XX OS
XX PN WO953051-A2.
XX PD
XX PD 21-OCT-1999.
XX PF 09-APR-1999; 99WO-1B000712.
XX PR 09-APR-1998; 98US-00057719.
XX PR 28-APR-1998; 98US-00069047.
XX PA (GEST ) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI; 2000-038446/03.
XX DR N-PSDB; AAZ43008.
XX PT Novel secreted protein 5' expressed sequence tag sequences used in
XX PT diagnostic, forensic, gene therapy, and chromosome mapping procedures.
XX PS Claim 3; Page 818; 837pp; English.
XX CC AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
XX CC sequences, corresponding to human secreted proteins. AAY64651 to AAY65438
XX CC represent the EST-related proteins corresponding to AAZ42265 to AAZ43052.
XX CC The 5' ESTs can be used for producing secreted human gene products. They
XX CC can be used to identify and isolate 5' untranslated regions (UTRs) and
XX CC upstream regulatory regions which control the location, development
XX CC stage, rate, and quantity of protein synthesis, as well as stability of
XX CC mRNA. The ESTs are also useful as probes for chromosome mapping, and to
XX CC obtain full length cDNA clones. The ESTs can also be used in forensic
XX CC procedures to identify individuals, or in diagnostic procedures to
XX CC identify individuals having genetic diseases resulting from abnormal gene
XX CC expression. The products may also be used in gene therapy protocols. The
XX CC nucleic acids encoding signal peptides can be used for directing
XX CC extracellular secretion of a polypeptide or the insertion of a
XX CC polypeptide into a membrane, or importing a polypeptide into a cell. The
XX CC proteins encoded by the EST sequences may be useful in treating a variety
XX CC of human conditions. Secreted proteins have therapeutic value, and the
XX CC identification of new secreted proteins is valuable. AAZ42249 to AAZ42264
XX CC and AAY64644 to AAY64650 represent sequences used in the exemplification
XX CC of the present invention
XX CC
XX CC Sequence 95 AA;
XX CC
XX CC Query Match 100.0%; Score 25; DB 3; Length 95;
XX CC Best Local Similarity 50.0%; Pred. No. 1.5e+03;
XX CC Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
XX CC
XX CC 1 XEFIXDX 8
XX CC :|||:|:|:
XX CC 47 QEFIDSDA 54
XX CC
XX CC RESULT 61
XX CC AAY92226
XX CC ID AAY92226 standard; protein; 95 AA.
XX CC
XX CC AAY92226;
XX CC 10-AUG-2000 (first entry)

```

XX DE Human endometrial specific steroid binding factor III.  
 XX XX  
 KW Endometrial specific steroid binding factor; ESBPIII; diagnosis;  
 KW gynaecological cancer; uterine; breast; endometrial; ovarian; antibody;  
 KW gene therapy; cytostatic.  
 XX XX  
 OS Homo sapiens.  
 XX XX  
 PN WO200020044-A1.  
 XX XX  
 PD 13-APR-2000.  
 XX XX  
 PF 30-SEP-1999; 99WO-US022753.  
 XX XX  
 PR 02-OCT-1998; 98US-0102743P.  
 XX XX  
 PA (DIAD-) DIADEXUS LLC.  
 XX XX  
 PI Macina RA;  
 XX XX  
 DR WPI; 2000-303649/26.  
 DR N-PSDB; AAA09104.  
 XX XX  
 PT Diagnosing, staging and monitoring gynaecological cancer comprising using  
 PT an elevated level of ESBPIII in a patient as an indicator of cancer.  
 XX XX  
 PS Claim 6; Page 28-29; 32pp; English.  
 XX XX  
 CC The levels of human endometrial specific steroid binding factor (ESBPIII)  
 CC can be measured and compared to control levels and used to diagnose the  
 CC presence of a gynaecological (uterine, breast, endometrial, or ovarian)  
 CC cancer in a patient. ESBPIII levels can also be used to diagnose  
 CC metastasis, to stage or monitor gynaecological cancer. Antibodies  
 CC specific for ESBPIII can be used to treat gynaecological cancers  
 XX XX  
 SQ Sequence 95 AA;  
 Query Match 100.0%; Score 25; DB 3; Length 95;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEFIXDX 8  
 Db :|||:|:  
 47 QEFIDSDA 54  
 RESULT 62  
 AAY92237  
 ID AAY92237 standard; protein; 95 AA.  
 AC AAY92237;  
 XX XX  
 DT 10-AUG-2000 (first entry)  
 XX XX  
 DE Mammoglobin homologue from clone Mamm-X.  
 XX XX  
 KW Clone Mamm-X; mammoglobin; breast cancer; cytostatic; anti-HIV;  
 KW immunosuppressive; anti-allergic; anti-infective; anti-inflammatory;  
 KW anti-arthritis; anti-arteriosclerotic; vasotropic; neuroprotective;  
 KW neutropic; dermatological; tranquilizer; vulnary.  
 XX XX  
 OS Homo sapiens.  
 XX XX  
 PN WO200020447-A2.  
 XX XX  
 PD 13-APR-2000.  
 XX XX  
 PF 06-OCT-1999; 99WO-US023294.  
 XX XX  
 PR 06-OCT-1998; 98US-0103195P.  
 PR 05-OCT-1999; 99US-00412231.  
 XX XX

PA (CURA-) CURAGEN CORP.  
 XX XX  
 PI Shinkets RA;  
 XX XX  
 DR WPI; 2000-303741/26.  
 DR N-PSDB; AAA09118.  
 XX XX  
 PT Nucleic acids encoding polypeptides with syncline-like, claudin-like or  
 PT cytokine-like activity, useful for treating diseases including cancer,  
 PT Alzheimer's and atherosclerosis.  
 XX XX  
 PS Claim 23; Fig 6; 118pp; English.  
 XX XX  
 CC Clone Mamm-X encodes a polypeptide that is 100 percent identical to human  
 CC Mammoglobin B precursor, a potential marker of breast cancer nodal  
 CC metastasis. The sequences are useful for treatment of diseases such as  
 CC cancer, immune disorders, autoimmune disease, acquired immune deficiency  
 CC syndrome (AIDS), transplant rejection, allergy, infection by a  
 CC pathological agent or organism, inflammatory disorders, arthritis, a  
 CC haematopoietic disorder, a skin disorder, atherosclerosis, restenosis, a  
 CC neurological disease, Alzheimer's disease, trauma, spinal cord injury and  
 CC skeletal disorders  
 XX XX  
 SQ Sequence 95 AA;  
 Query Match 100.0%; Score 25; DB 3; Length 95;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEFIXDX 8  
 Db :|||:|:  
 47 QEFIDSDA 54  
 RESULT 63  
 AAB31682  
 ID AAB31682 standard; protein; 95 AA.  
 AC AAB31682;  
 XX XX  
 DT 30-APR-2001 (first entry)  
 XX XX  
 DE An endometrial specific steroid binding factor III.  
 XX XX  
 KW Human; endometrial specific steroid binding factor; hESF; hESFI; hESFII;  
 KW hESFIII; inflammation; asthma; rhinitis; cystic fibrosis; airway disease;  
 KW neoplasia; atopy; phospholipase A2; polychlorinated biphenyl; chemotaxis;  
 KW phagocytosis; platelet aggregation; eicosanoid; endometrial cell.  
 XX XX  
 OS Homo sapiens.  
 XX XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT /note= "signal peptide"  
 XX XX  
 PN US6174992-B1.  
 XX XX  
 PD 16-JAN-2001.  
 XX XX  
 PF 08-MAR-1999; 99US-00263810.  
 XX XX  
 PR 21-MAR-1996; 96US-0014724P.  
 PR 21-MAR-1997; 97US-00821451.  
 XX XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX XX  
 PI Ni J, Yu G, Gentz R;  
 XX XX  
 DR WPI; 2001-158477/16.  
 DR N-PSDB; AAF25214.  
 XX XX  
 PT New human endometrial specific steroid binding factors, useful for  
 PT treating and preventing inflammation, asthma, rhinitis, cystic fibrosis,

PT airway disease, neoplasia and atopy.  
XX  
PS Claim 1; Fig 3; 36pp; English.  
XX  
CC The present sequence represents a human endometrial specific steroid  
CC binding factor (hESF). The specification describes hESFI, hESFII, and  
CC hESFIII. hESFI, II and III polypeptides, and polynucleotides encoding  
CC them are useful for treating and preventing inflammation, asthma,  
CC rhinitis, cystic fibrosis, airway disease, neoplasia and atopy.  
CC inhibiting phospholipase A2 activity, binding polychlorinated biphenyls,  
CC reducing foreign protein antigenicity, inhibiting monocyte and neutrophil  
CC chemotaxis and phagocytosis, inhibiting platelet aggregation, regulating  
CC eicosanoid levels in the human uterus, and for controlling the growth of  
CC endometrial cells. hESF polypeptides and nucleotides are also useful for  
CC research, biological, clinical or therapeutic purposes  
XX  
XX Sequence 95 AA;  
SQ  
Query Match 100.0%; Score 25; DB 4; Length 95;  
Best Local Similarity 50.0%; Pred. No. 1.5e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 1 XEFIXDX 8  
Db 47 QEFIDSDA 54  
RESULT 64  
AAO20555  
ID AAO20555 standard; protein; 95 AA.  
XX  
AC AAO20555;  
XX  
XX 27-JUN-2002 (first entry)  
XX  
XX Protein of human Lipophilin C.  
XX  
XX Immunogenic epitope; hormonally regulated organ; malignant tumour;  
XX Lipophilin; human.  
XX  
XX Homo sapiens.  
XX  
XX US2002034739-A1.  
XX  
XX 21-MAR-2002.  
XX  
XX 07-JUL-1998; 98US-00110716.  
XX  
XX 07-JUL-1998; 98US-00110716.  
XX  
XX (LEHR/) LEHRER R I.  
XX (ZHAO/) ZHAO C.  
XX (GLAS/) GLASGOW B J.  
XX  
XX Lehrer RI, Zhao C, Glasgow BJ;  
XX  
XX WPI; 2002-338922/37.  
XX N-PSDB; AAK99492.  
XX  
XX Peptides having the sequence of human lipophilin A, B and C are  
XX associated with carcinomas of hormonally regulated organs and are useful  
XX in the diagnosis and prognosis of various cancers.  
XX  
XX Claim 7; Fig 6; 22pp; English.  
XX  
XX The invention relates to a peptide comprising the amino acid sequences of  
XX human lipophilin A, B, or C or its allelic variant or fragment comprising  
XX at least one immunogenic epitope, which is purified and isolated, and may  
XX have the N-terminal acylated and/or C-terminal amidated or be a fusion  
XX protein. Molecules of the invention are used in the diagnosis and  
XX prognosis of malignant tumours associated with hormonally regulated  
XX organs such as uterus, ovary, prostate, testis, breast, kidney and  
XX thymus. This sequence represents the human Lipophilin C protein of the

CC invention  
XX  
XX Sequence 95 AA;  
SQ  
Query Match 100.0%; Score 25; DB 5; Length 95;  
Best Local Similarity 50.0%; Pred. No. 1.5e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 1 XEFIXDX 8  
Db 47 QEFIDSDA 54  
RESULT 65  
ABG96366  
ID ABG96366 standard; protein; 95 AA.  
XX  
AC ABG96366;  
XX  
XX 11-DEC-2002 (first entry)  
XX  
XX Human ovarian cancer marker M458.  
XX  
XX Human; ovarian cancer; marker; cancer; familial history; brain disorder;  
XX central nervous system disorder; bacterial meningitis; viral meningitis;  
XX Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;  
XX brain herniation; inflammation; encephalitis; testicular disorder;  
XX nontuberculous granulomatous orchitis; connective tissue disorder;  
XX heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;  
XX histological type; carcinogenic; ovarian cancer marker.  
XX  
XX Homo sapiens.  
XX  
XX WO200271928-A2.  
XX  
XX 19-SEP-2002.  
XX  
XX 14-MAR-2002; 2002WO-US007826.  
XX  
XX 14-MAR-2001; 2001US-0276025P.  
XX 14-MAR-2001; 2001US-0276026P.  
XX 10-AUG-2001; 2001US-0311732P.  
XX 19-SEP-2001; 2001US-0323580P.  
XX 26-SEP-2001; 2001US-0324967P.  
XX 26-SEP-2001; 2001US-0325102P.  
XX 26-SEP-2001; 2001US-0325149P.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;  
XX Meyers RE, Morrissey WP, Olandt PJ, Sen A, Vieby PO, Mills GB;  
XX Bast RC, Lu K, Schmandt RE, Zhao X, Giatt K;  
XX  
XX WPI; 2002-723277/78.  
XX N-PSDB; ABS76462.  
XX  
XX Assessing whether a patient is afflicted with ovarian cancer, useful in  
XX assessing the stage or progression of the disease, comprises comparing  
XX the expression level of a cancer marker in a sample from a patient and  
XX from a non cancer patient.  
XX  
XX Disclosure; Page 303; 481pp; English.  
XX  
XX The present invention relates to a new method for assessing whether a  
XX patient is afflicted with ovarian cancer. The method involves comparing  
XX the expression level of a marker in a patient sample and the normal level  
XX of expression of the marker in a control non-ovarian cancer sample, where  
XX the marker is selected from 363 cancer markers described in the  
XX specification. The method of the invention is useful in diagnosing or  
XX characterising cancer, in detecting the presence of cancer as early as  
XX possible, and the recurrence of ovarian cancer. The method may also be of  
XX particular use with patients having an enhanced risk of developing  
XX ovarian cancer (e.g. patients having a familial history of ovarian

CC cancer). The cancer markers may be used in the management and treatment  
 CC of e.g. brain and central nervous system disorders (e.g. bacterial and  
 CC viral meningitis, Alzheimer's disease or Parkinson's disease), brain  
 CC disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),  
 CC inflammations (e.g. bacterial or viral meningitis or encephalitis),  
 CC testicular disorders (e.g. bacterial or viral meningitis or encephalitis),  
 CC connective tissue disorders, or heart disorders (e.g. ischaemic heart  
 CC disease or atherosclerosis). The compositions and methods may also be  
 CC used in assessing the histological type of neoplasm associated with  
 CC ovarian cancer, monitoring the progression of ovarian cancer, determining  
 CC whether ovarian cancer has metastasized or is likely to metastasize,  
 CC selecting a composition for inhibiting ovarian cancer, assessing the  
 CC ovarian carcinogenic potential of a compound, or inhibiting ovarian  
 CC cancer or at risk of developing ovarian cancer. The present amino acid  
 CC sequence represents one of the ovarian cancer markers described in the  
 CC invention

XX Sequence 95 AA;

Query Match 100.0%; Score 25; DB 5; Length 95;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEPIXXDX 8  
 :|||:::  
 Db 47 QEFIDSDA 54

RESULT 66

ABB09635

ID ABB09635 standard; protein; 95 AA.

XX AC ABB09635;

XX 29-MAY-2002 (first entry)

XX Human endometrial specific steroid-binding factor (hESF) III.

XX Human; endometrial specific steroid-binding factor; hESF;

XX prostatic steroid-binding protein; hESF I; hESF II; hESF III; asthma.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1. .21

XX Protein /note= "signal peptide"

XX /note= "mature protein"

XX US6338948-B1.

XX 15-JAN-2002.

XX 30-MAY-2000; 2000US-00583169.

XX 21-MAR-1996; 96US-0014724P.

XX 21-MAR-1997; 97US-00821451.

XX 08-MAR-1999; 99US-00263810.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ni J, Yu G, Gentz R;

XX WPI; 2002-215019/27.

XX N-PSDB; ABL41783.

XX New antibody specific for human endometrial specific steroid-binding  
 PT factor (hESF) III, useful for detecting hESF III protein in biological  
 PT sample and to isolate or identify clones expressing the protein.

XX Disclosure; Fig 3; 36pp; English.

XX The present sequence represents a endometrial specific steroid-binding

CC factor (hESF) III. The full length protein has a molecular weight of 8.10  
 CC kDa. The protein has homology to rat prostatic steroid-binding protein  
 CC C3. Antibodies which bind hESF proteins, such as hESF I, hESF II, and  
 CC hESF III are useful for isolating or to identify clones expressing the  
 CC polypeptides or to purify the polypeptides by affinity chromatography.  
 CC Agonists and antagonists of hESF proteins are useful for treating and/or  
 CC preventing susceptibility to asthma

XX Sequence 95 AA;

Query Match 100.0%; Score 25; DB 5; Length 95;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEPIXXDX 8  
 :|||:::  
 Db 47 QEFIDSDA 54

RESULT 67

ABG73169

ID ABG73169 standard; protein; 95 AA.

XX AC ABG73169;

XX 08-APR-2003 (first entry)

XX Human endometrial specific steroid-binding factor III (hESF III) protein.

XX Human; endometrial specific steroid-binding factor; hESF;

XX Clara cell 10 kDa; CC10; secretory protein; asthma;

XX prostatic steroid-binding protein; hormone; lung; uterus; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1. .21

XX Protein /label= Signal\_peptide

XX /label= Mature\_hESF\_III

XX US2002151012-A1.

XX 17-OCT-2002.

XX 06-NOV-2001; 2001US-00985911.

XX 21-MAR-1996; 96US-0014724P.

XX 21-MAR-1997; 97US-00821451.

XX 08-MAR-1999; 99US-00263810.

XX 30-MAY-2000; 2000US-00583169.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ni J, Yu G, Gentz R;

XX WPI; 2003-182506/18.

XX N-PSDB; ABX14917.

XX New human endometrial specific steroid-binding factor (hESF) proteins and  
 PT genes, useful for treating or diagnosing a disease or susceptibility to a  
 PT disease, particularly asthma.

XX Claim 15; Fig 3; 37pp; English.

XX The invention discloses isolated polypeptides, which comprise human  
 CC endometrial specific steroid-binding factors I, II and III (hESF I, II  
 CC and III), and the nucleic acids encoding them. The hESF polypeptide has  
 CC homologues to mammalian Clara cell 10 kDa (CC10) secretory protein and  
 CC rat prostatic steroid-binding protein which are factors which modulate or  
 CC mediate the action of hormones involved in the regulation of functions of  
 CC the lung and uterus. The nucleic acids and polypeptides can be used to  
 CC identify compounds that bind to and inhibit activation, raise antibodies

CC or develop antagonists against the isolated hESF polypeptide. The  
CC polypeptides or polynucleotides are useful for treating a patient having  
CC a need of hESF I, hESF III or for treating a patient having a  
CC need to inhibit hESF. The polypeptide is administered by providing to the  
CC patient the DNA encoding the hESF polypeptide in vivo (gene therapy). In  
CC particular, the disease is asthma. The hESF polypeptides or  
CC polynucleotides are also useful for diagnosing a disease or a  
CC susceptibility to the disease. The sequence presented is the hESF III  
CC protein  
XX  
SQ Sequence 95 AA;

Query Match 100.0%; Score 25; DB 6; Length 95;  
Best Local Similarity 50.0%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db 47 QEFIDSDA 54

RESULT 68  
ABR47522  
ID ABR47522 standard; protein; 95 AA.

XX ABR47522;

DT 12-JUN-2003 (first entry)

DE Breast cancer associated protein sequence SEQ ID NO:280.

XX Human; breast cancer; cytostatic; gene therapy.

OS Homo sapiens.

XX WO2003004989-A2.

PD 16-JAN-2003.

XX 21-JUN-2002; 2002WO-US019669.

PR 21-JUN-2001; 2001US-0299887P.

PR 27-JUN-2001; 2001US-0301572P.

PR 18-JUL-2001; 2001US-0306501P.

PR 23-SEP-2001; 2001US-0325002P.

PR 05-MAR-2002; 2002US-0362585P.

PR 14-MAY-2002; 2002US-0380391P.

XX (MILL-) MILLENIUM PHARM INC.

XX Lillie J, Gannavarapu M, Glatt K, Hoersh S, Kamatkar S;

PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;

PI Bast RC, Hortobagyi GN, Puztai L, Meric F, Sahin A, Mills GB;

XX WPI; 2003-210381/20.

DR N-PSDB; ACC50218.

XX Breast cancer diagnosis or treatment by comparing the level of expression

PT of a marker in a patient sample with that in the control non-breast

PT cancer sample.

XX Claim 1; SEQ ID NO 280; 128pp; English.

CC The present invention describes a method for assessing whether a patient

CC is afflicted with breast cancer. The method comprises comparing the level

CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and

CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 95 AA;

Query Match 100.0%; Score 25; DB 6; Length 95;  
Best Local Similarity 50.0%; Pred. No. 1.5e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db 47 QEFIDSDA 54

RESULT 69

ADC78795

ID ADC78795 standard; protein; 95 AA.

XX ADC78795;

DT 01-JAN-2004 (first entry)

DE Human PRO protein #12.

XX human; PRO; cancer; inflammatory bowel disease; ulcerative colitis;

XX Crohn's disease.

OS Homo sapiens.

XX WO2003034984-A2.

PD 01-MAY-2003.

XX 15-OCT-2002; 2002WO-US033070.

XX 19-OCT-2001; 2001US-0340083P.

XX (GETH ) GENENTECH INC.

XX Goddard A, Gurney AL;

XX WPI; 2003-481990/45.

XX N-PSDB; ADC78794.

XX New PRO polynucleotide and polypeptide, useful for the manufacture of a

PT medicament for diagnosing or treating cancer or inflammatory bowel

PT disorder e.g., ulcerative colitis or Crohn's disease.

XX Claim 12; SEQ ID NO 24; 327pp; English.

XX The invention comprises the amino acid and coding sequences of human PRO

CC proteins. The DNA and protein sequences of the invention are useful for

CC the diagnosis and treatment of cancer and inflammatory bowel disease

CC (e.g. ulcerative colitis or Crohn's disease). The present amino acid

CC sequence represents a human PRO protein of the invention.

XX  
SQ Sequence 95 AA;

Query Match 100.0%; Score 25; DB 7; Length 95;

Best Local Similarity 50.0%; Pred. No. 1.5e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8

Db 47 QEFIDSDA 54

RESULT 70

ABP01586

ID ABP01586 standard; protein; 97 AA.

XX ABP01586;

XX AC ABP01586;

```
XX 25-JUN-2002 (first entry)
XX Human ORFX protein sequence SEQ ID NO:3154.
XX
XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
XX Homo sapiens.
XX
XX WO200192523-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US010836.
XX
XX 30-MAY-2000; 2000US-0206132P.
XX
XX 29-AUG-2000; 2000US-0226716P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach MD;
XX
XX WPI; 2002-106308/14.
XX
XX N-PSDB; ABN17338.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX
XX Disclosure; SEQ ID NO 3154; 1037pp; English.
XX
XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification)). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP0010 to ABP1500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 97 AA;
XX
XX Query Match 100.0%; Score 25; DB 5; Length 97;
XX Best Local Similarity 50.0%; Pred. No. 1.6e+03;
XX Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 XEFIXDXD 8
XX :|||:
XX Db 53 IEFISEDE 60
XX
XX Query Match 100.0%; Score 25; DB 5; Length 97;
XX Best Local Similarity 50.0%; Pred. No. 1.6e+03;
XX Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 XEFIXDXD 8
XX :|||:
XX Db 53 IEFISEDE 60
XX
```

```
RESULT 71
ABB63376
ID ABB63376 standard; protein; 98 AA.
XX
XX ABB63376;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 16920.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL07479.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Disclosure; SEQ ID NO 16920; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC AB372072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 98 AA;
XX
XX Query Match 100.0%; Score 25; DB 4; Length 98;
XX Best Local Similarity 50.0%; Pred. No. 1.6e+03;
XX Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 XEFIXDXD 8
XX :|||:
XX Db 46 LEFINSDC 53
XX
XX RESULT 72
XX AAE03588
XX ID AAE03588 standard; protein; 98 AA.
XX
XX AAE03588;
XX
XX 06-AUG-2001 (first entry)
XX
XX Wheat aspartyl-tRNA synthetase from clone wleln.pk0021.e6.
XX
XX Wheat; aspartyl-tRNA synthetase; herbicide; genetic mapping;
KW plant breeding.
XX
```

```

OS Triticum aestivum.
XX
PN US6255090-B1.
XX
XX 03-JUL-2001.
XX
XX 14-JUL-1999; 99US-00352990.
XX
XX 15-JUL-1998; 98US-0092866P.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
PA (PION-) PIONEER HI-BRED INT INC.
XX
XX Famodu LO, Orozco EM, Rafalski JA;
XX WPI; 2001-388927/41.
DR N-PSDB; AAD07972.
DR
XX
XX New isolated polynucleotide encoding an aspartyl-tRNA synthetase useful
PT as targets to facilitate design and/or identification of inhibitors of
PT those enzymes that may be useful as herbicides.
XX
XX Claim 12; Col 33-36; 40pp; English.
XX
XX The present sequence is wheat aspartyl-tRNA synthetase of the invention.
CC The aspartyl-tRNA synthetase are used as targets to facilitate designing
CC and identification of inhibitors of the enzymes which are useful as
CC herbicides. All or a substantial portion of the nucleic acid fragments of
CC the present invention are used as probes for genetically and physically
CC mapping the genes that they are a part of, and as markers for traits
CC linked to those genes. Such information is useful in plant breeding in
CC order to develop lines with desired phenotypes
XX
XX Sequence 98 AA;
XX
Query Match 100.0%; Score 25; DB 4; Length 98;
Best Local Similarity 50.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db :|||:|:
74 CEFIGLDA 81
RESULT 73
AAG34364
ID AAG34364 standard; protein; 99 AA.
XX
AC AAG34364;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 41798.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.

```





XX The invention relates to isolated Acinetobacter baumannii nucleic acids.  
 CC The A. baumannii nucleic acids and polypeptides are useful as reagents  
 CC for diagnosing a bacterial disease, as components of antibacterial  
 CC vaccines, as targets for antibacterial drugs, to detect the presence of  
 CC A. baumannii and other Acinetobacter species in a sample, in screening  
 CC compounds for the ability to interfere with the A. baumannii life cycle  
 CC or to inhibit A. baumannii infection, and as biocontrol agents for  
 CC plants. The present sequence represents the amino acid sequence of an A.  
 CC baumannii protein.  
 XX  
 CC  
 SQ Sequence 99 AA;

Query Match 100.0%; Score 25; DB 6; Length 99;  
 Best Local Similarity 50.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDXD 8  
 :|||:|:  
 Db 38 PEFIVKDK 45

RESULT 75  
 ID AAW72393  
 AA AAW72393 standard; protein; 101 AA.  
 XX  
 AC AAW72393;  
 XX  
 DT 02-FEB-1999 (first entry)  
 XX  
 DE Pathogen response protein LSD1-interacting protein OO.  
 XX  
 KW LSD1-interacting protein OO; plant pathogen response; apoptosis;  
 KW programmed cell death; disease resistance; herbicide resistance;  
 KW transgenic plant; crop protection.  
 XX  
 OS Arabidopsis thaliana.

XX  
 XX  
 XX W09837755-A1.  
 XX  
 PD 03-SEP-1998.  
 XX  
 XX 27-FEB-1998; 98WO-US004077.  
 XX  
 XX 28-FEB-1997; 97US-0039063P.  
 XX  
 XX (UYNC-) UNIV NORTH CAROLINA.  
 XX  
 XX Dangl JL, Dietrich RA, Richberg MH, Eppe PM;  
 PI  
 XX WPI; 1998-531501/45.  
 XX  
 DR N-PSDB; AAV66764.  
 XX  
 PT New isolated Arabidopsis genes - useful for producing transgenic plants  
 PT which show resistance to cell death caused by pathogens or herbicides.  
 PT  
 XX  
 XX Claim 46; Page 65; 88pp; English.

XX This is the amino acid sequence of LSD1-interacting protein OO of  
 CC Arabidopsis thaliana. LSD1 interacting genes (see AAV66755-67) were  
 CC isolated from a yeast gene expression library constructed in plasmid pJG4  
 CC -5 using RNA from Arabidopsis leaves infected with Pseudomonas syringae.  
 CC A two-hybrid system was used with LSD1 short and long open reading frames  
 CC (see AAV66750-51) as bait. LSD1 (see AAW72366-67) is a novel polypeptide  
 CC that regulates the initial response of plants to pathogens and the  
 CC subsequent spread of plant cell death engendered by infection. Since the  
 CC inactivation of LSD1 by mutation leads to enhanced disease resistance,  
 CC LSD1 partner proteins represent novel targets for engineering plants with  
 CC enhanced resistance to pathogens. Thus, the invention includes all  
 CC proteins (see AAW72384-96) that interact with the cell death regulator  
 CC LSD1  
 XX  
 XX Sequence 101 AA;

Query Match 100.0%; Score 25; DB 2; Length 101;  
 Best Local Similarity 50.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEFIXDXD 8  
 :|||:|:  
 Db 75 IEFICLDL 82

RESULT 76  
 ID ABP32702  
 AA ABP32702 standard; protein; 103 AA.  
 XX  
 AC ABP32702;  
 XX  
 DT 08-JUL-2002 (first entry)  
 XX  
 DE Human ORF1675 protein, SEQ ID NO:3350.  
 XX  
 KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KW immune modulation; haematopoiesis regulation; tissue growth;  
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 KW behaviour; cancer; proliferative disorder; neurological disorder;  
 KW cardiovascular disease; immune system disorder; organ transplantation;  
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
 KW vasotrophic; antipsoriatic; antidiabetic; cytostatic; nootropic;  
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W0200190366-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 XX 24-MAY-2001; 2001WO-US017076.  
 XX  
 XX 24-MAY-2000; 2000US-0206690P.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Leach MD, Shinkets RA;  
 PI  
 XX WPI; 2002-106200/14.  
 XX  
 DR N-PSDB; ABN76728.  
 XX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation.  
 PT  
 XX  
 XX Claim 10; Page 1096; 2508pp; English.

XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,

CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antiinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases  
 XX  
 SQ Sequence 103 AA;

Query Match 100.0%; Score 25; DB 5; Length 103;  
 Best Local Similarity 50.0%; Pred. No. 1.7e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||:::  
 Db 29 GEFIXGDX 36

RESULT 77  
 ADA54452  
 ID ADA54452 standard; protein; 105 AA.

XX ADA54452;

XX 20-NOV-2003 (first entry)

XX Human protein, SEQ ID 2020.

XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
 KW Gene therapy; human; secretory protein; membrane proteins; cancer;  
 KW inflammatory disease; osteoporosis; neurological disease.

XX Homo sapiens.

XX EP1293569-A2.

XX 19-MAR-2003.

XX 21-MAR-2002; 2002EP-00006586.

XX 14-SEP-2001; 2001JP-00328381.

XX 24-JUN-2002; 2002US-0350433P.

XX (HELI-) HELIX RES INST.

PA (REAS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

XX WPI; 2003-395539/38.

DR N-PSDB; ADA52813.

XX New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.

PS Claim 14; SEQ ID NO 2020; 205pp; English.

XX The present invention relates to novel human secretory or membrane

CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
 CC ADA54071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.  
 XX  
 SQ Sequence 105 AA;

Query Match 100.0%; Score 25; DB 6; Length 105;  
 Best Local Similarity 50.0%; Pred. No. 1.7e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||:::  
 Db 4 EEFIFDY 11

RESULT 78  
 ABB17370  
 ID ABB17370 standard; protein; 106 AA.

XX ABB17370;

XX 23-JAN-2002 (first entry)

XX Human nervous system related polypeptide SEQ ID NO 6027.

XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
 KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.

XX Homo sapiens.

XX WO200159063-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US001334.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

```
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 03-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.

PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX N-PSDB; ABA13696.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
XX useful for preventing, diagnosing and/or treating nervous system cancers
XX and metastases.
XX
XX Claim 11; SEQ ID NO 6027; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins
XX (AB114678-AB18001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
XX breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
XX cardiovascular disorders such as myocardial ischaemias; (d) wound healing
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pt_sequences
XX
XX Sequence 106 AA;
XX
XX Query Match 100.0%; Score 25; DB 4; Length 106;
XX Best Local Similarity 50.0%; Pred.No. 1.7e+03;
XX Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 XEFIXDX 8
XX :|||:|:
XX Db 21 AEFILSDP 28
```

RESULT 79  
ABUS8228  
ID ABUS8228 standard; protein; 106 AA.  
XX  
XX AC ABUS8228;  
XX  
XX DT 14-APR-2003 (first entry)  
XX  
XX DE Wheat stress response protein #18.  
XX  
XX KW Plant; EST; expressed sequence tag; stress response; drought; heat;  
XX radiation; pathogen attack; grain flavour; disease resistance;  
XX KW peptide-methionine sulfoxide reductase; DNA repair; enzyme;  
XX intracellular protein transport.  
XX  
XX OS Triticum aestivum.  
XX  
XX PN US2002152497-A1.  
XX  
XX PD 17-OCT-2002.  
XX  
XX PF 19-FEB-2002; 2002US-00078929.  
XX  
XX PR 07-MAY-1999; 99US-0133038P.  
XX PR 07-MAY-1999; 99US-0133042P.  
XX PR 11-MAY-1999; 99US-0133427P.  
XX PR 11-MAY-1999; 99US-0133428P.  
XX PR 11-MAY-1999; 99US-0133436P.  
XX PR 11-MAY-1999; 99US-0133437P.  
XX PR 11-MAY-1999; 99US-0133438P.  
XX PR 04-JUN-1999; 99US-0137667P.  
XX PR 05-MAY-2000; 2000US-00566394.  
XX  
XX PA (FALC/) FALCO S C.  
XX PA (FAMO/) FAMODU O O.  
XX PA (MEYE/) MEYERS B C.  
XX PA (MIAO/) MIAO G.  
XX PA (ODEL/) ODELL J T.  
XX PA (RAFA/) RAFALSKI J A.  
XX PA (THOR/) THORPE C J.  
XX PA (SAKA/) SAKAI H.  
XX PA (WENG/) WENG Z.  
XX  
XX PI Falco SC, Famodu OO, Meyers BC, Miao G, Odell JT, Rafalski JA,  
XX Thorpe CJ, Sakai H, Weng Z;  
XX  
XX DR WPI; 2003-198391/19.  
XX DR N-PSDB; ABX78399.  
XX  
XX PT New peptide-methionine sulfoxide reductase and nucleic acids, useful in  
XX improving plant response to stress, engineering plants with increased  
XX disease and stress resistance, or and improving/protecting grain flavor.  
XX  
XX PS Claim 1; Page 134; 205pp; English.  
XX  
XX CC The invention relates to isolated nucleic acids encoding plant stress  
XX response proteins (including peptide-methionine sulfoxide reductases)  
XX appearing as ABUS8148-ABUS8246 (or a protein 80% identical to them) from  
XX Zea mays, Oryza sativa, Glycine max, or Triticum aestivum. Also included  
XX are expression cassettes, transformed host cells, transgenic  
XX plants/seeds, modulating the level of peptide-methionine sulfoxide  
XX reductase in a plant and a computer system/data processing system for  
XX identifying, analysing, or modelling a genetic sequence. The plant  
XX nucleic acid is useful in developing strategies to improve plant response  
XX to stress (e.g. drought, heat, radiation or pathogen attack), engineering  
XX plants with increased disease and stress resistance, manipulating DNA  
XX repair and recombination efficiency, manipulating intracellular protein  
XX transport, and improving/protecting grain flavour. The nucleic acids may  
XX also be used as probes or amplification primers in the detection,  
XX quantitation or isolation of gene transcripts, for recombinant expression  
XX of encoded polypeptides, as immunogens in preparing or screening

CC antibodies, and in sense or antisense suppression of one or more genes in  
CC a host cell, tissue or plant. The proteins may be used as immunogens or  
CC antigens to obtain antibodies specifically immunoreactive with the  
CC protein, and in assays for enzyme agonists or antagonists. The present  
CC sequence is a plant stress response protein (or fragment)  
XX  
XX SQ Sequence 106 AA;

Query Match 100.0%; Score 25; DB 6; Length 106;  
Best Local Similarity 50.0%; Pred. NO. 1.7e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIIXDX 8  
:|:|:|:  
Db 71 SEFIKIDA 78

## RESULT 80

ADC97504  
ID ADC97504 standard; protein; 106 AA.  
XX  
XX AC ADC97504;  
XX  
XX DT 01-JAN-2004 (first entry)  
XX  
XX DE E. faecium protein sequence SEQ ID 7131.  
XX  
XX KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;  
XX abdominal-pelvic infection.  
XX  
XX OS Enterococcus faecium.  
XX  
XX PN US6583275-B1.  
XX  
XX PD 24-JUN-2003.  
XX  
XX PF 30-JUN-1998; 98US-00107532.  
XX  
XX PR 02-JUL-1997; 97US-0051571P.  
XX PR 14-MAY-1998; 98US-0085598P.  
XX  
XX PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX PI Doucette-Stamm LA, Bush D;  
XX  
XX DR WPI; 2003-799836/75.  
XX DR N-PSDB; ADC93850.  
XX  
XX PT New isolated nucleic acid derived from Enterococcus faecium encoding an  
XX Enterococcus faecium polypeptide useful for detection, prevention and  
XX treatment of a pathological condition resulting from a bacterial  
XX infection.  
XX  
XX PS Example 1; SEQ ID NO 7131; 243pp; English.  
XX  
XX CC The invention relates to an isolated nucleic acid derived from  
XX Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
XX one of 10 fully defined sequences given in the (or comprising 40  
XX sequential nucleotides chosen from any of the nucleic acids, its  
XX complement or sequences hybridising to it). Also included are a  
XX recombinant vector comprising the nucleic acid operably linked to  
XX transcription regulatory element, a cell comprising the vector and a  
XX single-stranded probe comprising the nucleic acid. The nucleic acids are  
XX chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
XX The nucleic acids is useful for diagnosing pathological conditions  
XX resulting from E. faecium bacterial infection (e.g. urinary tract  
XX infection, bacteraemia, endocarditis, wounds and abdominal-pelvic  
XX infection) and for screening drugs such as agonists and antagonists. The  
XX nucleic acid is useful for recombinant production of Candida albicans -  
XX derived peptides or antisense polypeptides. Pharmaceutical compositions  
XX and vaccines containing the nucleic acid are useful for preventing or  
XX treating Enterococcus faecium infections. The present sequence represents  
XX one if the disclosed E. faecium proteins.





```
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159394P.
PR 13-OCT-1999; 99US-0159395P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159594P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 26-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 110;
Best Local Similarity 50.08; Pred. No. 1.8e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX B
Db 84 IEFICLDL 91

RESULT 84
AAG24527
ID AAG24527 standard; protein; 110 AA.
XX
AC AAG24527;
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 28236.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW
```

hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-00301439.

25-FEB-1999; 99US-0121825P.

05-MAR-1999; 99US-0123180P.

09-MAR-1999; 99US-0123548P.

23-MAR-1999; 99US-0125788P.

23-MAR-1999; 99US-0126264P.

29-MAR-1999; 99US-0126785P.

01-APR-1999; 99US-0127462P.

06-APR-1999; 99US-0128234P.

08-APR-1999; 99US-0128714P.

16-APR-1999; 99US-0129845P.

19-APR-1999; 99US-0130077P.

21-APR-1999; 99US-0130449P.

23-APR-1999; 99US-0130510P.

23-APR-1999; 99US-0130891P.

28-APR-1999; 99US-0131449P.

30-APR-1999; 99US-0132048P.

30-APR-1999; 99US-0132407P.

04-MAY-1999; 99US-0132484P.

05-MAY-1999; 99US-0132485P.

06-MAY-1999; 99US-0132486P.

06-MAY-1999; 99US-0132487P.

07-MAY-1999; 99US-0132863P.

11-MAY-1999; 99US-0134256P.

14-MAY-1999; 99US-0134218P.

14-MAY-1999; 99US-0134219P.

14-MAY-1999; 99US-0134221P.

14-MAY-1999; 99US-0134370P.

18-MAY-1999; 99US-0134768P.

19-MAY-1999; 99US-0134941P.

20-MAY-1999; 99US-0135124P.

21-MAY-1999; 99US-0135353P.

24-MAY-1999; 99US-0135629P.

25-MAY-1999; 99US-0136021P.

27-MAY-1999; 99US-0136392P.

28-MAY-1999; 99US-0136782P.

01-JUN-1999; 99US-0137222P.

03-JUN-1999; 99US-0137528P.

04-JUN-1999; 99US-0137502P.

07-JUN-1999; 99US-0137724P.

08-JUN-1999; 99US-0138094P.

10-JUN-1999; 99US-0138540P.

10-JUN-1999; 99US-0138847P.

14-JUN-1999; 99US-0139119P.

16-JUN-1999; 99US-0139452P.

16-JUN-1999; 99US-0139453P.

17-JUN-1999; 99US-0139492P.

18-JUN-1999; 99US-0139454P.

18-JUN-1999; 99US-0139455P.

18-JUN-1999; 99US-0139456P.

18-JUN-1999; 99US-0139457P.

18-JUN-1999; 99US-0139458P.

18-JUN-1999; 99US-0139459P.

18-JUN-1999; 99US-0139460P.

18-JUN-1999; 99US-0139461P.

18-JUN-1999; 99US-0139462P.

18-JUN-1999; 99US-0139463P.

18-JUN-1999; 99US-0139750P.

18-JUN-1999; 99US-0139763P.

21-JUN-1999; 99US-0139817P.

22-JUN-1999; 99US-0139899P.

23-JUN-1999; 99US-0140353P.

23-JUN-1999; 99US-0140354P.

PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145244P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147182P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.

PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155113P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 110;  
Best Local Similarity 50.0%; Pred. No. 1.8e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFTXDX 8  
Db :|||:|:  
84 IEFICLDL 91

RESULT 85

ABR57093  
ID ABR57093 standard; peptide; 110 AA.

XX ABR57093;

XX 26-AUG-2003 (first entry)

XX MLHR comparison related HU.IGEREC amino acid sequence.

XX LHR; MLHR; immunoglobulin; Ig; immunoglobulin heavy chain dimer;  
KW ligand-combined partner; cell surface adhesion molecule; ligand;  
KW lymphocyte cell surface glycoprotein.

OS Unidentified.

XX JP2002325589-A.

XX





PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130443P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131443P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140961P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142154P.  
PR 06-JUL-1999; 99US-0142055P.  
PR 08-JUL-1999; 99US-0142300P.  
PR 09-JUL-1999; 99US-0142803P.  
PR 12-JUL-1999; 99US-0142920P.  
PR 13-JUL-1999; 99US-0142972P.  
PR 14-JUL-1999; 99US-0143542P.  
PR 15-JUL-1999; 99US-0143624P.  
PR 16-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.

PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 31-AUG-1999; 99US-0151303P.  
PR 01-SEP-1999; 99US-0151438P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158212P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.

```

PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 114;
Best Local Similarity 50.0%; Pred. No. 1.8e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 105 IEFICLDL 112

RESULT 88
ABB54747
ID ABB54747 standard; protein; 114 AA.
XX AC ABB54747;
XX 29-AUG-2003 (revised)
DT 16-MAY-2002 (first entry)
XX DE Lactococcus lactis protein p1334.
XX KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX OS Lactococcus lactis; ILL403.
XX PN FR2807446-A1.
XX PD 12-OCT-2001.
XX PF 11-APR-2000; 2000FR-00004630.
XX PR 11-APR-2000; 2000FR-00004630.
XX PS (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX WIPI; 2002-043418/06.
XX New nucleotide sequence useful in the identification of Lactococcus
XX lactis and related species.
XX Claim 6; SEQ ID NO 1449; 2504pp; French.
XX The present invention is related to a Lactococcus lactis nucleotide
XX sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic
XX acid sequence is useful in the detection and/or amplification of nucleic
CC

PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 114;
Best Local Similarity 50.0%; Pred. No. 1.8e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 105 IEFICLDL 112

RESULT 88
ABB54747
ID ABB54747 standard; protein; 114 AA.
XX AC ABB54747;
XX 29-AUG-2003 (revised)
DT 16-MAY-2002 (first entry)
XX DE Lactococcus lactis protein p1334.
XX KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX OS Lactococcus lactis; ILL403.
XX PN FR2807446-A1.
XX PD 12-OCT-2001.
XX PF 11-APR-2000; 2000FR-00004630.
XX PR 11-APR-2000; 2000FR-00004630.
XX PS (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX WIPI; 2002-043418/06.
XX New nucleotide sequence useful in the identification of Lactococcus
XX lactis and related species.
XX Claim 6; SEQ ID NO 1449; 2504pp; French.
XX The present invention is related to a Lactococcus lactis nucleotide
XX sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic
XX acid sequence is useful in the detection and/or amplification of nucleic
CC

CC acid sequence, particularly to identify Lactococcus lactis or related
CC species. The proteins of the invention are useful for the biosynthesis or
CC biodegradation of a composition of interest. The invention helps research
CC in lactic bacteria, particularly useful in the production of yogurt and
CC cheese. Note: The sequence data for this patent is based on equivalent
CC patent WO200177334 (published 18-OCT-2001) which is available in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
XX Sequence 114 AA;
Query Match 100.0%; Score 25; DB 5; Length 114;
Best Local Similarity 50.0%; Pred. No. 1.8e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 22 GEFIVCDE 29

RESULT 89
AAU02032
ID AAU02032 standard; protein; 116 AA.
XX AC AAU02032;
XX 29-AUG-2001 (first entry)
DT DE B. thuringiensis toxic crystal protein, CryET56.
XX KW Delta endotoxin; Lepidopteran-active; crystal protein; insecticide;
XX transgenic plant; corn; wheat; soybean; oat; cotton; rice; sorghum;
XX sugarcane; tomato; tobacco; kapok; flax; potato; barley; turf grass;
XX pasture grass; berry; fruit; legume; vegetable; ornamental plant; shrub;
XX cactus; tree cell; gypsy moth; looper; tobacco budworm; spruce budworm;
XX cotton leaf perforator; CryET56.
XX OS Bacillus thuringiensis.
XX PN WO200119859-A2.
XX PD 22-MAR-2001.
XX PF 13-SEP-2000; 2000WO-US025361.
XX PR 15-SEP-1999; 99US-0153995P.
XX PA (MONS ) MONSANTO CO.
XX PI Baum JA, Chu C, Donovan WP, Gilmer AJ, Rupar MJ;
XX WIPI; 2001-281518/29.
XX N-PSDB; AAS02475.
XX Lepidopteran-active Bacillus thuringiensis delta-endotoxin polypeptides
XX and the polynucleotides that encode them, useful for increasing the
XX insect resistance of plant.
XX Claim 5; Page 121; 173pp; English.
XX The sequence represents a B. thuringiensis Lepidopteran-active delta-
XX endotoxin, crystal protein CryET56. The Lepidopteran-active B.
XX thuringiensis delta-endotoxin polypeptides may be used as compositions
XX that are applied to plant crops to protect them from insect damage. The
XX polynucleotides may be used in the production of transgenic plants that
XX express the insecticidal polypeptides and consequently have improved
XX insect resistance compared to non-transformed plants. Monocotyledonous or
XX dicotyledonous plants may be protected in this way, for example corn,
XX wheat, soybean, oat, cotton, rice, rye, sorghum, sugarcane, tomato,
XX tobacco, kapok, flax, potato, barley, turf grass, pasture grass, berry,
XX fruit, legume, vegetable, ornamental plant, shrub, cactus and/or tree
XX cell. A wide range of insects (e.g. gypsy moth, looper, tobacco budworm,

```

CC cotton leaf perforator and spruce budworm) may be affected by application  
CC of the insecticidal polypeptides (full details given in specification)  
XX

SQ Sequence 116 AA;

Query Match 100.0%; Score 25; DB 4; Length 116;  
Best Local Similarity 50.0%; Pred. No. 1.9e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 XEFIXDX 8  
:||||:|:  
Db 95 FEFIPVDA 102

RESULT 90  
ABU15149  
ID ABU15149 standard; protein; 116 AA.

XX AC ABU15149;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #676.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Escherichia coli.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA19019.

XX PS Claim 25; SEQ ID NO 43073; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 116 AA;

Query Match 100.0%; Score 25; DB 6; Length 116;  
Best Local Similarity 50.0%; Pred. No. 1.9e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 XEFIXDX 8

Db 11 PEFIFNDN 18

RESULT 91

ABU11397  
ID ABU11397 standard; protein; 120 AA.

XX AC ABU11397;

XX DT 11-FEB-2003 (first entry)

XX DE Protein encoded by *S. atroolivaceus* leinamycin gene cluster ORF lnmV.

XX KW Leinamycin biosynthesis gene cluster; lnm; open reading frame; ORF;

XX KW anti-tumour antibiotic; broad spectrum antimicrobial activity;

XX KW Gram-positive; Gram-negative bacteria; chemical modification; metabolite;

XX KW apo-carrier protein; holo-carrier protein; tumour; polypeptide;

XX KW hybrid polypeptide/polypeptide metabolite; lnm production; cytostatic.

XX OS Streptomyces atroolivaceus.

XX PN WO200277179-A2.

XX PD 03-OCT-2002.

XX PF 22-MAR-2002; 2002WO-US008937.

XX PR 26-MAR-2001; 2001US-0278935P.

XX PA (REGC ) UNIV CALIFORNIA.

XX PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX PI Shen B, Cheng Y, Tang G;

XX DR WPI; 2003-018907/01.

XX DR N-PSDB; ABX34289.

XX PT Novel gene cluster responsible for synthesis of leinamycin in

XX PT Streptomyces atroolivaceus useful for making various peptide and/or

XX PS polypeptide, and/or hybrid polypeptide/polypeptide metabolites.

XX CC Claim 13; Page 152; 185pp; English.

XX CC The present invention relates to the isolation of the Streptomyces  
XX CC atroolivaceus leinamycin (lnm) biosynthesis gene cluster containing 71  
XX CC open reading frames (ORFs) (ORFs -35 through -1, ORFs lnmA through lnmZ,  
XX CC and ORFs +1 through +9). Leinamycin is a novel anti-tumour antibiotic  
XX CC produced by several Streptomyces species. It exhibits broad spectrum  
XX CC antimicrobial activity against Gram-positive and Gram-negative bacteria,  
XX CC but not against fungi. The polypeptides encoded by the lnm biosynthesis  
XX CC gene cluster ORFs are useful for chemically modifying a molecule in a  
XX CC host cell. The host cell is a bacterium or eukaryotic cell, including a

CC mammalian, yeast, plant, fungal, or insect cell. The molecule is an  
 CC endogenous metabolite produced by the host cell or exogenously supplied  
 CC metabolite, or an amino acid, and the polypeptide is a peptide synthetase  
 CC or amino transferase. The polypeptides encoded by the lmm gene cluster  
 CC are useful for converting an apo-carrier protein to a holo-carrier  
 CC protein. lmm shows potent antitumor activity in tumour models in vivo.  
 CC The lmm gene cluster modules and/or catalytic domains are useful for  
 CC making various peptide and/or polyketide, and/or hybrid  
 CC polypeptide/polyketide metabolites. The proteins encoded by the ORFs are  
 CC useful alone, or in combination with other active domains to modify  
 CC various target substrates. The lmm gene cluster is useful to upregulate  
 CC endogenous lmm production to permit lmm production in cells and/or to  
 CC make various modified lmm. lmm, its analogue, or other polyketide,  
 CC peptide or hybrid polyketide/peptide metabolites are useful as  
 CC therapeutic agents, to treat a number of disorders, depending upon the  
 CC type of metabolites. ABU1341-ABU11411 represent the proteins encoded by  
 CC ORFs of the *S. atroolivaceus* leinamycin biosynthesis gene cluster  
 XX  
 SQ Sequence 120 AA;

Query Match 100.0%; Score 25; DB 6; Length 120;

Best Local Similarity 50.0%; Pred. NO. 1.9e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIGXDX 8  
 Db :|||::|:  
 69 NEFIQTDD 76

## RESULT 92

AAG27381

ID AAG27381 standard; protein; 121 AA.

XX AAG27381;

AC AAG27381;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 32195.

XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX Arabidopsis thaliana.

XX  
 PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 05-MAR-1999; 99US-0123180P.

XX 09-MAR-1999; 99US-0123548P.

XX 23-MAR-1999; 99US-0125788P.

XX 25-MAR-1999; 99US-0126264P.

XX 29-MAR-1999; 99US-0126785P.

XX 01-APR-1999; 99US-0127462P.

XX 06-APR-1999; 99US-0128234P.

XX 08-APR-1999; 99US-0128714P.

XX 16-APR-1999; 99US-0129845P.

XX 19-APR-1999; 99US-0130077P.

XX 21-APR-1999; 99US-0130449P.

XX 23-APR-1999; 99US-0130510P.

XX 23-APR-1999; 99US-0130891P.

XX 28-APR-1999; 99US-0131449P.

XX 30-APR-1999; 99US-0132048P.

XX 30-APR-1999; 99US-0132407P.

XX 04-MAY-1999; 99US-0132484P.

XX 05-MAY-1999; 99US-0132485P.

XX 06-MAY-1999; 99US-0132486P.

XX 06-MAY-1999; 99US-0132487P.

XX 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.  
 PR 14-MAY-1999; 99US-0134218P.  
 PR 14-MAY-1999; 99US-0134219P.  
 PR 14-MAY-1999; 99US-0134221P.  
 PR 14-MAY-1999; 99US-0134370P.  
 PR 18-MAY-1999; 99US-0134768P.  
 PR 19-MAY-1999; 99US-0134941P.  
 PR 20-MAY-1999; 99US-0135124P.  
 PR 21-MAY-1999; 99US-0135353P.  
 PR 24-MAY-1999; 99US-0135629P.  
 PR 25-MAY-1999; 99US-0136021P.  
 PR 27-MAY-1999; 99US-0136392P.  
 PR 28-MAY-1999; 99US-0136782P.  
 PR 01-JUN-1999; 99US-0137222P.  
 PR 03-JUN-1999; 99US-0137528P.  
 PR 04-JUN-1999; 99US-0137502P.  
 PR 07-JUN-1999; 99US-0137724P.  
 PR 08-JUN-1999; 99US-0138094P.  
 PR 10-JUN-1999; 99US-0138540P.  
 PR 10-JUN-1999; 99US-0138847P.  
 PR 14-JUN-1999; 99US-0139119P.  
 PR 16-JUN-1999; 99US-0139452P.  
 PR 16-JUN-1999; 99US-0139453P.  
 PR 17-JUN-1999; 99US-0139492P.  
 PR 18-JUN-1999; 99US-0139454P.  
 PR 18-JUN-1999; 99US-0139455P.  
 PR 18-JUN-1999; 99US-0139456P.  
 PR 18-JUN-1999; 99US-0139457P.  
 PR 18-JUN-1999; 99US-0139458P.  
 PR 18-JUN-1999; 99US-0139459P.  
 PR 18-JUN-1999; 99US-0139460P.  
 PR 18-JUN-1999; 99US-0139461P.  
 PR 18-JUN-1999; 99US-0139462P.  
 PR 18-JUN-1999; 99US-0139463P.  
 PR 18-JUN-1999; 99US-0139750P.  
 PR 21-JUN-1999; 99US-0139763P.  
 PR 21-JUN-1999; 99US-0139817P.  
 PR 22-JUN-1999; 99US-0139899P.  
 PR 23-JUN-1999; 99US-0140353P.  
 PR 23-JUN-1999; 99US-0140354P.  
 PR 24-JUN-1999; 99US-0140695P.  
 PR 28-JUN-1999; 99US-0140823P.  
 PR 29-JUN-1999; 99US-0140991P.  
 PR 30-JUN-1999; 99US-0141287P.  
 PR 01-JUL-1999; 99US-0141842P.  
 PR 01-JUL-1999; 99US-0142154P.  
 PR 02-JUL-1999; 99US-0142055P.  
 PR 06-JUL-1999; 99US-0142390P.  
 PR 08-JUL-1999; 99US-0142803P.  
 PR 09-JUL-1999; 99US-0142920P.  
 PR 12-JUL-1999; 99US-0142977P.  
 PR 13-JUL-1999; 99US-0143542P.  
 PR 14-JUL-1999; 99US-0143624P.  
 PR 15-JUL-1999; 99US-0144005P.  
 PR 16-JUL-1999; 99US-0144085P.  
 PR 16-JUL-1999; 99US-0144086P.  
 PR 19-JUL-1999; 99US-0144325P.  
 PR 19-JUL-1999; 99US-0144331P.  
 PR 19-JUL-1999; 99US-0144332P.  
 PR 19-JUL-1999; 99US-0144333P.  
 PR 19-JUL-1999; 99US-0144334P.  
 PR 19-JUL-1999; 99US-0144335P.  
 PR 20-JUL-1999; 99US-0144352P.  
 PR 20-JUL-1999; 99US-0144632P.  
 PR 20-JUL-1999; 99US-0144884P.  
 PR 21-JUL-1999; 99US-0144814P.  
 PR 21-JUL-1999; 99US-0145086P.  
 PR 21-JUL-1999; 99US-0145088P.  
 PR 22-JUL-1999; 99US-0145085P.  
 PR 22-JUL-1999; 99US-0145087P.  
 PR 22-JUL-1999; 99US-0145089P.  
 PR 22-JUL-1999; 99US-0145192P.  
 PR 23-JUL-1999; 99US-0145145P.

PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0156659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159283P.  
PR 13-OCT-1999; 99US-0159284P.  
PR 13-OCT-1999; 99US-0159285P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.

PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161040P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 121;

Best Local Similarity 50.0%; Pred. No. 2e+03; 0; Indels 0; Gaps 0;

Matches 4; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 XEPIXDX 8

Db 104 LERIEDE 111

RESULT 93

ABB81063

ID ABB81063 standard; protein; 123 AA.

AC ABB81063;

XX 05-NOV-2002 (first entry)

DT XX

DE XX Canine low affinity IGE receptor (CD23) fragment PsCaCD23\_123.

XX Canine; immunoglobulin E; IGE; CD23; antiallergic; antiasthmatic;

KW antiinflammatory; dermatological; gene therapy; vaccine; receptor.

XX OS

XX US6410714-B1.

XX 25-JUN-2002.

XX 24-MAR-2000; 2000US-00535521.

XX 24-MAR-1999; 99US-0125913P.

XX (HESK-) HESKA CORP.

XX Weber ER, McCall CA;

XX WPI; 2002-588896/63.

XX N-PSDB; ABN86575.

XX New isolated canine low affinity immunoglobulin E receptor nucleic acid

PT molecule, useful for protecting canids from diseases mediated by the

PT receptor, such as allergy, atopic dermatitis, asthma, and hay fever.

XX Claim 7; Col 61-62; 33pp; English.

XX The invention relates to isolated canine low affinity immunoglobulin E

CC (IGE) receptor (CD23) polypeptides and encoding nucleic acid molecules.

CC The CD23 polypeptides can be expressed by standard recombinant

CC methodology. The CD23 polynucleotides are useful for protecting canids

CC from diseases mediated by CD23, for developing compounds that regulate

CC IGE and/or CD23 levels in a canid for treating allergy related diseases

CC such as atopic dermatitis, asthma, hay fever and food sensitivities. The

CC present sequence represents a CD23 receptor PsCaCD23\_123

XX SQ Sequence 123 AA;

Query Match

Best Local Similarity 100.0%; Score 25; DB 5; Length 123;

Matches 4; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```
Qv 1 XEFIXDX 8
Db :|||:|:|:
73 GEFIWDE 80

RESULT 94
ID AAG01456
XX AAG01456 standard; protein; 124 AA.
AC AAG01456;
XX 06-OCT-2000 (first entry)
XX Human secreted protein, SEQ ID NO: 5537.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX Homo sapiens.
XX EPI033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-00200610.
XX 26-FEB-1999; 99US-0122487P.
XX (GEST ) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX N-PSDB; AAC01462.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX Claim 13; SEQ ID NO 5537; 71pp + Sequence Listing; English.
XX The present sequence is a polypeptide encoded by one of a large number of
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC tissues. EST sequences usually correspond mainly to the 3' untranslated
CC region (UTR) of the mRNA because they are often obtained from oligo-dT
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC are also used in diagnostic, forensic, gene therapy and chromosome
CC mapping procedures. They are used to obtain upstream regulatory sequences
CC and to design expression and secretion vectors
XX Sequence 124 AA;
XX Query Match 100.0%; Score 25; DB 3; Length 124;
XX Best Local Similarity 50.0%; Pred. No. 2e+03;
XX Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Qv 1 XEFIXDX 8
Db :|||:|:|:
113 EEFIFYD 120

RESULT 95
ID AAG34531
XX AAG34531 standard; protein; 124 AA.
XX AAG34531;
XX 18-OCT-2000 (first entry)
DT Arabidopsis thaliana protein fragment SEQ ID NO: 42032.
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EPI033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0132048P.
XX 04-MAY-1999; 99US-0132407P.
XX 05-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 07-MAY-1999; 99US-0132487P.
XX 11-MAY-1999; 99US-0132863P.
XX 14-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 14-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134768P.
XX 19-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135353P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.
XX 03-JUN-1999; 99US-0137528P.
XX 07-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.
XX 10-JUN-1999; 99US-0138540P.
XX 10-JUN-1999; 99US-0138847P.
XX 14-JUN-1999; 99US-0139119P.
XX 16-JUN-1999; 99US-0139452P.
XX 16-JUN-1999; 99US-0139453P.
XX 17-JUN-1999; 99US-0139492P.
XX 18-JUN-1999; 99US-0139454P.
XX 18-JUN-1999; 99US-0139455P.
XX 18-JUN-1999; 99US-0139456P.
XX 18-JUN-1999; 99US-0139457P.
XX 18-JUN-1999; 99US-0139458P.
XX 18-JUN-1999; 99US-0139459P.
XX 18-JUN-1999; 99US-0139460P.
XX 18-JUN-1999; 99US-0139461P.
XX 18-JUN-1999; 99US-0139462P.
XX 18-JUN-1999; 99US-0139463P.
XX 18-JUN-1999; 99US-0139750P.
XX 18-JUN-1999; 99US-0139763P.
```

PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145115P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151063P.  
PR 27-AUG-1999; 99US-0151066P.

PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160983P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 124;

Best Local Similarity 50.0%; Pred. No. 2e+03; Mismatches 4; Indels 0; Gaps 0;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db 46 QEFIVDDL 53

RESULT 96

AAU16559

ID AAU16559 standard; protein; 125 AA.

XX AAU16559;

AC AAU16559;

XX 07-NOV-2001 (first entry)

DT Human novel secreted protein, Seq ID 1512.

DE Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;





PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI  
 XX  
 XX WPI; 2001-488783/53.  
 DR N-PSDB; AAS26546.  
 XX  
 XX  
 PT New nucleic acid molecules encoding 461 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives.  
 XX  
 XX  
 PS Claim 11; SEQ ID NO 1512; 980pp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
 CC in diagnosing a pathological condition or susceptibility to a  
 CC alleviating symptoms associated with the disorders and in diagnostic  
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,  
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
 CC and many other disorders listed in the specification. The polypeptides  
 CC can also be used to aid wound healing and epithelial cell proliferation,  
 CC to prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 Query Match 100.0%; Score 25; DB 4; Length 125;  
 Best Local Similarity 50.0%; Pred. No. 2e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 XEFIXDX 8  
 Db 25 AEFIKRDD 32  
 RESULT 97  
 ABUS55628  
 ID ABUS55628 standard; protein; 125 AA.  
 XX  
 XX ABUS55628;  
 AC  
 XX  
 XX 18-MAR-2003 (first entry)  
 DT  
 XX  
 XX Human novel polypeptide #715.  
 DE  
 XX  
 XX Human; neural disorder; immune system disorder; renal disorder;  
 KW muscular disorder; respiratory disease; reproductive disorder;  
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;  
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
 KW haemostatic; antiarteriosclerotic.  
 XX  
 XX Homo sapiens.  
 OS  
 XX US2002132753-A1.  
 PN

XX  
 PD 19-SEP-2002.  
 XX  
 PF 17-JAN-2001; 2001US-00764864.  
 XX  
 XX 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 XX  
 XX (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 XX  
 PI Rosen CA, Ruben SM, Barash SC;  
 XX  
 XX WPI; 2003-147444/14.  
 DR N-PSDB; ABX73887.  
 XX  
 PT New polypeptides and nucleic acids, useful in gene therapy for treating,  
 PT inhibiting or preventing e.g. neural, immune system, muscular,  
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
 PT renal disorders.  
 XX  
 XX Claim 11; SEQ ID NO 1512; 402pp; English.  
 PS  
 XX The invention relates to human novel polypeptides and their associated  
 CC

CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
 CC therapy for treating, inhibiting or preventing neural disorders, immune  
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left  
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and  
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood  
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
 CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and  
 CC ABU55748 represent human novel polypeptides of the invention  
 XX  
 SQ Sequence 125 AA;

Query Match 100.0%; Score 25; DB 6; Length 125;

Best Local Similarity 50.0%; Pred. No. 2e+03;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXXX 8

:|:|:|:|:

Db 25 AEFIKRDD 32

RESULT 98

AAG06674

ID AAG06674 standard; protein; 126 AA.

XX AC AAG06674;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 3533.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX FN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 08-APR-1999; 99US-0128234P.

PR 16-APR-1999; 99US-0128714P.

PR 19-APR-1999; 99US-0129845P.

PR 21-APR-1999; 99US-0130077P.

PR 23-APR-1999; 99US-0130449P.

PR 28-APR-1999; 99US-0130810P.

PR 30-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 04-MAY-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 07-MAY-1999; 99US-0132487P.

PR 11-MAY-1999; 99US-0132863P.

PR 14-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 14-MAY-1999; 99US-0134370P.

PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.

PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.

PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137528P.

PR 04-JUN-1999; 99US-0137502P.

PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.

PR 10-JUN-1999; 99US-0138847P.

PR 14-JUN-1999; 99US-0139119P.

PR 16-JUN-1999; 99US-0139452P.

PR 16-JUN-1999; 99US-0139453P.

PR 17-JUN-1999; 99US-0139492P.

PR 18-JUN-1999; 99US-0139454P.

PR 18-JUN-1999; 99US-0139455P.

PR 18-JUN-1999; 99US-0139456P.

PR 18-JUN-1999; 99US-0139457P.

PR 18-JUN-1999; 99US-0139458P.

PR 18-JUN-1999; 99US-0139459P.

PR 18-JUN-1999; 99US-0139460P.

PR 18-JUN-1999; 99US-0139461P.

PR 18-JUN-1999; 99US-0139462P.

PR 18-JUN-1999; 99US-0139463P.

PR 18-JUN-1999; 99US-0139750P.

PR 18-JUN-1999; 99US-0139763P.

PR 21-JUN-1999; 99US-0139817P.

PR 22-JUN-1999; 99US-0139899P.

PR 23-JUN-1999; 99US-0140353P.

PR 23-JUN-1999; 99US-0140354P.

PR 24-JUN-1999; 99US-0140895P.

PR 28-JUN-1999; 99US-0140823P.

PR 29-JUN-1999; 99US-0140991P.

PR 30-JUN-1999; 99US-0141287P.

PR 01-JUL-1999; 99US-0141842P.

PR 01-JUL-1999; 99US-0142154P.

PR 02-JUL-1999; 99US-0142055P.

PR 06-JUL-1999; 99US-0142390P.

PR 08-JUL-1999; 99US-0142803P.

PR 09-JUL-1999; 99US-0142920P.

PR 12-JUL-1999; 99US-0142977P.

PR 13-JUL-1999; 99US-0143542P.

PR 14-JUL-1999; 99US-0143624P.

PR 15-JUL-1999; 99US-0144005P.

PR 16-JUL-1999; 99US-0144085P.

PR 16-JUL-1999; 99US-0144086P.

PR 19-JUL-1999; 99US-0144325P.

PR 19-JUL-1999; 99US-0144331P.

PR 19-JUL-1999; 99US-0144332P.

PR 19-JUL-1999; 99US-0144333P.

PR 19-JUL-1999; 99US-0144334P.

PR 20-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.

PR 20-JUL-1999; 99US-0144632P.

PR 20-JUL-1999; 99US-0144884P.

PR 21-JUL-1999; 99US-0144814P.

PR 21-JUL-1999; 99US-0145086P.

PR 21-JUL-1999; 99US-0145088P.

PR 22-JUL-1999; 99US-0145085P.

PR 22-JUL-1999; 99US-0145087P.

PR 22-JUL-1999; 99US-0145089P.

PR 22-JUL-1999; 99US-0145192P.

PR 23-JUL-1999; 99US-0145145P.

PR 23-JUL-1999; 99US-0145218P.

PR 23-JUL-1999; 99US-0145224P.

PR 26-JUL-1999; 99US-0145276P.

```
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147933P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153707P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155655P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 06-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158023P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159322P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 21-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160818P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.

PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 126;
Best Local Similarity 50.0%; Pred. No. 2e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 46 QEFIVDDL 53

RESULT 99
AAG35079
ID AAG35079 standard; protein; 126 AA.
XX
AC AAG35079;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 42798.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123184P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
```

PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140921P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142820P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.

PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 100.0%; Score 25; DB 3; Length 126;  
Best Local Similarity 50.0%; Pred. No. 2e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:|:  
Db 46 QEFIVDDL 53

## RESULT 100

ABG03470  
ID ABG03470 standard; protein; 126 AA.

XX AC ABG03470;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #3461.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN W0200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-ESDB; AAS67657.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity.

XX PS Claim 20; SEQ ID NO 33829; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain  
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX CC and in recombinant production of (II). The polynucleotides are also used  
XX CC in diagnostics as expressed sequence tags for identifying expressed  
XX CC genes. (I) is useful in gene therapy techniques to restore normal  
XX CC activity of (II) or to treat disease states involving (II). (II) is  
XX CC useful for generating antibodies against it, detecting or quantitating a  
XX CC polypeptide in tissue, as molecular weight markers and as a food  
XX CC supplement. (II) and its binding partners are useful in medical imaging  
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders  
XX CC involving aberrant protein expression or biological activity. The  
XX CC polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
XX CC amino acid sequences of the invention. Note: The sequence data for this  
XX CC patent did not appear in the printed specification, but was obtained in  
XX CC electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 126 AA;

XX SQ

Query Match 100.0%; Score 25; DB 4; Length 126;  
Best Local Similarity 50.0%; Pred. No. 2e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:|:  
Db 56 IEFIFDI 63

Search completed: September 12, 2004, 02:34:00  
Job time : 111 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2004, 01:55:47 ; Search time 49 Seconds

(without alignments)  
15.705 Million cell updates/sec

Title: US-09-660-302E-1

Perfect score: 25

Sequence: 1 XEPIXXDX 8

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

PIR\_78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	38	S68260	hypothetical prote
2	25	100.0	54	E45881	hypothetical 5.9K
3	25	100.0	67	E81922	hypothetical prote
4	25	100.0	68	C96900	small conserved pr
5	25	100.0	79	G97292	hypothetical prote
6	25	100.0	91	W5WL13	B5 protein - human
7	25	100.0	98	S62338	L71-10 protein - f
8	25	100.0	101	S12170	hypothetical prote
9	25	100.0	109	T10321	hypothetical prote
10	25	100.0	114	G86800	prophage p13 prote
11	25	100.0	116	A85636	hypothetical prote
12	25	100.0	118	Q0B8BH	hypothetical 13K p
13	25	100.0	119	F82462	conserved hypochet
14	25	100.0	122	D97300	probable membrane
15	25	100.0	123	AC3470	hypothetical prote
16	25	100.0	124	T48833	hypothetical prote
17	25	100.0	129	JT0666	hypothetical 14.7K
18	25	100.0	135	AH1093	hypothetical prote
19	25	100.0	140	T08492	probable polypepti
20	25	100.0	149	G97212	probable acetyltra
21	25	100.0	151	D81333	probable protein-t
22	25	100.0	151	T28645	transcription fact
23	25	100.0	154	A60998	replication protei
24	25	100.0	154	A87354	hypothetical prote
25	25	100.0	155	B96920	uncharacterized co
26	25	100.0	157	A61755	hypothetical prote
27	25	100.0	160	AG1752	bacteriophage prot
28	25	100.0	160	C82083	hypothetical prote
29	25	100.0	163	D83883	hypothetical prote

30	25	100.0	164	G81321	probable integral
31	25	100.0	165	G82910	peptide methionine
32	25	100.0	167	AE3385	acetyltransferase
33	25	100.0	170	G95987	probable ureidogly
34	25	100.0	170	G97106	uncharacterized co
35	25	100.0	171	E71530	hypothetical prote
36	25	100.0	171	C84141	hypothetical prote
37	25	100.0	172	CFYCBB	C-phycocyanin beta
38	25	100.0	174	C82053	shikimate kinase V
39	25	100.0	176	G90357	dtpp-4-dehydrotham
40	25	100.0	176	D97084	enzyme of dihydrof
41	25	100.0	177	H97205	probable kinase fr
42	25	100.0	181	A64393	hypothetical prote
43	25	100.0	181	T48241	hypothetical prote
44	25	100.0	181	H72630	hypothetical prote
45	25	100.0	183	C83913	hypothetical prote
46	25	100.0	185	S78329	photosystem I chai
47	25	100.0	186	E69176	hypothetical prote
48	25	100.0	187	G97069	precorrin-6B methy
49	25	100.0	189	C97705	hypothetical prote
50	25	100.0	190	A70100	hypothetical prote
51	25	100.0	192	E47700	glycerol metabolis
52	25	100.0	192	AD3014	conserved hypochet
53	25	100.0	194	E84829	hypothetical prote
54	25	100.0	196	S64583	hypothetical prote
55	25	100.0	196	D90869	probable regulator
56	25	100.0	196	G85519	probable regulator
57	25	100.0	196	F64755	ykGK protein - Esc
58	25	100.0	199	T02787	probable submergen
59	25	100.0	201	S73365	translation initia
60	25	100.0	202	T50322	hypothetical coile
61	25	100.0	202	AB1527	hypothetical prote
62	25	100.0	202	AH1169	hypothetical prote
63	25	100.0	203	AC1595	hypothetical prote
64	25	100.0	204	F64353	dCTP deaminase (EC
65	25	100.0	204	G83060	probable ribosomal
66	25	100.0	204	A28864	proteinase B (EC 3
67	25	100.0	205	C98270	hypothetical prote
68	25	100.0	207	H71671	ribosomal protein
69	25	100.0	207	E97825	50S ribosomal prot
70	25	100.0	208	F95892	probable glutathio
71	25	100.0	208	D82532	50S ribosomal prot
72	25	100.0	208	T17329	hypothetical prote
73	25	100.0	209	I40484	uracil phosphoribo
74	25	100.0	209	C71647	hypothetical prote
75	25	100.0	209	H97864	hypothetical prote
76	25	100.0	209	E89009	protein R08E5.1 [i
77	25	100.0	210	T41553	thymidylate kinase
78	25	100.0	210	S28955	dTMP kinase (EC 2.
79	25	100.0	211	A85098	hypothetical prote
80	25	100.0	211	C84775	probable harpin-in
81	25	100.0	211	JC4883	scytalidopepsin B
82	25	100.0	212	A83521	conserved hypochet
83	25	100.0	213	E82675	hypothetical prote
84	25	100.0	214	A70198	general stress pro
85	25	100.0	214	E69339	conserved hypochet
86	25	100.0	217	T07595	glutathione transf
87	25	100.0	218	T01076	transcription fact
88	25	100.0	218	D90179	hypothetical prote
89	25	100.0	219	A69505	uridyate kinase (
90	25	100.0	221	H97021	hypothetical prote
91	25	100.0	222	G82361	methyltransferase-
92	25	100.0	222	B90913	hypothetical prote
93	25	100.0	223	AF1345	hypothetical prote
94	25	100.0	225	WMNV29	AcOrf-16 DA26 prot
95	25	100.0	226	T23611	hypothetical prote
96	25	100.0	226	AB1477	hypothetical prote
97	25	100.0	226	A11115	hypothetical prote
98	25	100.0	226	D90057	hypothetical prote
99	25	100.0	227	C71246	hypothetical prote
100	25	100.0	227	C75210	sterol biosynthesi
101	25	100.0	227	B84667	hypothetical prote
102	25	100.0	229	T41762	BV/ODV-E26 orf16 -

103	25	100.0	229	2	G72337	hypothetical prote	176	25	100.0	290	2	E71631	lipid A biosynthes
104	25	100.0	231	2	F70713	hypothetical prote	177	25	100.0	291	2	H89894	conserved hypothet
105	25	100.0	231	2	T48386	hypothetical prote	178	25	100.0	291	2	AH1308	methyltransferases
106	25	100.0	232	2	H90214	hypothetical prote	179	25	100.0	291	2	AH1680	glucose-1-phosphat
107	25	100.0	232	2	B83873	transcription regu	180	25	100.0	292	2	S23342	glucose-1-phosphat
108	25	100.0	232	2	G72259	probable di-trans,	181	25	100.0	292	2	S15301	ABC transporter (A
109	25	100.0	233	2	T10198	hypothetical prote	182	25	100.0	292	2	AH1354	glucose-1-phosphat
110	25	100.0	233	1	B46350	25K protein - whit	183	25	100.0	293	2	F64969	formylmethanofuran
111	25	100.0	235	1	E75375	conserved hypothet	184	25	100.0	296	2	S57647	uncharacterized pr
112	25	100.0	236	2	S01086	hypothetical prote	185	25	100.0	297	2	C97192	transcription regu
113	25	100.0	236	2	S01086	conserved hypothet	186	25	100.0	297	2	AF3651	hypothetical prote
114	25	100.0	240	1	B64457	uridylylate kinase h	187	25	100.0	297	2	T29255	hypothetical prote
115	25	100.0	240	1	F64676	biotin synthase p	188	25	100.0	297	2	AC2380	adhesin homolog HI
116	25	100.0	240	2	AD1715	hypothetical prote	189	25	100.0	298	2	I64138	hypothetical prote
117	25	100.0	240	2	AD1715	hypothetical prote	190	25	100.0	298	2	S69523	hypothetical prote
118	25	100.0	240	2	AD1715	hypothetical prote	191	25	100.0	298	2	H70483	probable glycosylt
119	25	100.0	243	2	C84551	probable ribosomal	192	25	100.0	298	2	H97186	transcription regu
120	25	100.0	244	2	B70422	hypothetical prote	193	25	100.0	302	2	AH1749	conserved hypothet
121	25	100.0	244	2	T15807	hypothetical prote	194	25	100.0	305	2	C86745	pantothenate kinas
122	25	100.0	245	2	G30282	hypothetical prote	195	25	100.0	306	2	D86805	hypothetical prote
123	25	100.0	248	2	F97246	probable membrane	196	25	100.0	306	2	S60906	htrp protein - Pse
124	25	100.0	250	1	OSNC2	cytochrome-c oxida	197	25	100.0	308	2	C86498	AcCoA carboxylase/
125	25	100.0	250	2	S05065	hyperosmolarity-re	198	25	100.0	308	2	C72124	acetyl-coenzyme A
126	25	100.0	250	2	AF1727	hypothetical prote	199	25	100.0	308	2	E38447	oligopeptide trans
127	25	100.0	251	2	G81279	enterochelin uptak	200	25	100.0	308	2	H87306	glycosyl transfera
128	25	100.0	252	1	S05629	cytochrome-c oxida	201	25	100.0	309	2	T25800	C2H2-type zinc fin
129	25	100.0	254	1	H89454	enoyl-CoA hydratase	202	25	100.0	311	2	D84116	ribose ABC transpo
130	25	100.0	255	2	B70116	hypothetical prote	203	25	100.0	312	2	F90339	drpP-Glucose 4,6-d
131	25	100.0	255	2	A87619	hypothetical prote	204	25	100.0	312	2	A87449	conserved hypothet
132	25	100.0	255	2	G30403	conserved hypothet	205	25	100.0	313	2	F89866	hypothetical prote
133	25	100.0	256	2	B69930	hypothetical prote	206	25	100.0	313	2	T03049	hypothetical prote
134	25	100.0	256	2	C90763	hypothetical prote	207	25	100.0	313	2	T29691	hypothetical prote
135	25	100.0	256	2	D85626	hypothetical prote	208	25	100.0	315	2	G82242	conserved hypothet
136	25	100.0	257	2	D90246	diphthine synthase	209	25	100.0	316	1	LUJF12	annexin XII Hydr
137	25	100.0	258	2	C69475	conserved hypothet	210	25	100.0	317	2	H70306	RNA polymerase alp
138	25	100.0	258	2	S74631	hypothetical prote	211	25	100.0	317	2	A75008	udp-glucose 4-epim
139	25	100.0	260	1	JXEBT	lysine/arginine/or	212	25	100.0	318	2	H71145	probable UDP-gluc
140	25	100.0	262	2	T48480	ribosomal protein	213	25	100.0	318	2	T33577	hypothetical prote
141	25	100.0	262	2	S65594	ferrichrome-iron r	214	25	100.0	318	2	D87506	glycosyl transfera
142	25	100.0	262	2	G91294	ferric iron reduct	215	25	100.0	321	1	LNHUR	IgB FC receptor II
143	25	100.0	262	2	B86136	ferric iron reduct	216	25	100.0	322	2	AD1719	oligopeptide ABC t
144	25	100.0	262	2	T02115	hypothetical prote	217	25	100.0	322	2	AH1348	oligopeptide ABC t
145	25	100.0	263	2	A83961	ribonuclease III r	218	25	100.0	322	2	D97345	oligopeptide ABC t
146	25	100.0	263	2	A83503	2-hydroxymuconic s	219	25	100.0	322	2	A97766	cell filamentation
147	25	100.0	265	2	AH3312	hypothetical cytos	220	25	100.0	323	2	F71130	probable oligopept
148	25	100.0	266	2	C75179	hypothetical prote	221	25	100.0	323	2	E97291	oligopeptide ABC t
149	25	100.0	266	2	G71034	hypothetical prote	222	25	100.0	323	2	C83940	sugar ABC transpor
150	25	100.0	266	2	G89796	conserved hypothet	223	25	100.0	323	2	D85074	hypothetical prote
151	25	100.0	270	1	JX0064	interleukin-1 alph	224	25	100.0	323	2	S59373	cyclin homolog UME
152	25	100.0	270	1	C9ECID	cell division inhi	225	25	100.0	324	2	T17978	probable ribonucle
153	25	100.0	270	2	E90837	cell division inhi	226	25	100.0	325	2	T16995	probable cinnamyl-
154	25	100.0	270	2	F85695	cell division inhi	227	25	100.0	326	2	G75117	dipeptide abc tran
155	25	100.0	270	2	F85695	septum site determ	228	25	100.0	326	2	B69412	transcription int
156	25	100.0	270	2	AH0724	probable senescenc	229	25	100.0	326	2	AD3647	bactoprenol glucos
157	25	100.0	270	2	E84578	protoporphyrinogen	230	25	100.0	326	2	H97346	oligopeptide ABC t
158	25	100.0	271	1	ICHU1A	interleukin-1 alph	231	25	100.0	327	1	RDVZAS	ribonucleoside-dip
159	25	100.0	273	2	A69001	conserved hypothet	232	25	100.0	327	2	B69856	oligopeptide trans
160	25	100.0	275	2	T32626	hypothetical prote	233	25	100.0	328	2	S78498	probable glucose-1
161	25	100.0	276	2	T52526	monophenol monooxy	234	25	100.0	329	1	A36952	CDP-6-deoxy-Delta(
162	25	100.0	278	2	A56421	casein kinase II (	235	25	100.0	329	2	AE0378	cdp-6-deoxy-Delta-
163	25	100.0	281	2	S34496	hypothetical prote	236	25	100.0	329	2	F69051	3-isopropylmalate
164	25	100.0	282	2	T49722	related to host-sp	237	25	100.0	329	2	A86768	GMP reductase (EC
165	25	100.0	282	2	G97709	hypothetical prote	238	25	100.0	329	2	G71014	probable dipeptide
166	25	100.0	282	2	S74644	regulatory compone	239	25	100.0	330	2	JQ2374	hemagglutinin - in
167	25	100.0	284	1	A41382	UTP-glucose-1-phos	240	25	100.0	330	2	JQ2375	hemagglutinin - in
168	25	100.0	284	2	T17595	hypothetical prote	241	25	100.0	330	2	C90090	ATP (GTP)-binding p
169	25	100.0	286	2	E97241	hypothetical prote	242	25	100.0	331	2	JQ2377	hemagglutinin - in
170	25	100.0	288	2	B81240	glucose-1-phosphat	243	25	100.0	331	2	JQ2378	hemagglutinin - in
171	25	100.0	288	2	H82014	glucose-1-phosphat	244	25	100.0	331	2	G75099	sugar transferase
172	25	100.0	289	2	B86649	hypothetical prote	245	25	100.0	333	2	A12131	hypothetical prote
173	25	100.0	290	2	B61267	arylamine N-acetyl	246	25	100.0	336	2	H82040	general secretion
174	25	100.0	290	2	S41536	glucose-1-phosphat	247	25	100.0	336	2	F75540	hypothetical prote
175	25	100.0	290	2	S59322	hypothetical prote	248	25	100.0	336	2	A97168	glycosyltransferas



249 25 100.0 337 1 PAPGF  
250 25 100.0 337 2 S70469  
251 25 100.0 337 2 S70469  
252 25 100.0 338 2 AE0917  
253 25 100.0 338 2 AE0917  
254 25 100.0 339 2 A46666  
255 25 100.0 339 2 E86761  
256 25 100.0 340 2 S62493  
257 25 100.0 340 2 T44330  
258 25 100.0 341 2 H90267  
259 25 100.0 342 2 A45462  
260 25 100.0 342 2 B97330  
261 25 100.0 346 2 B71651  
262 25 100.0 346 2 F81351  
263 25 100.0 346 2 D81288  
264 25 100.0 347 1 MNXWE  
265 25 100.0 347 2 A72618  
266 25 100.0 347 2 S52173  
267 25 100.0 347 2 S52174  
268 25 100.0 347 2 S52175  
269 25 100.0 347 2 S52176  
270 25 100.0 347 2 S52178  
271 25 100.0 347 2 S52179  
272 25 100.0 347 2 S52183  
273 25 100.0 347 2 S52187  
274 25 100.0 347 2 S52188  
275 25 100.0 347 2 S52191  
276 25 100.0 347 2 S52195  
277 25 100.0 347 2 S52197  
278 25 100.0 347 2 S52198  
279 25 100.0 347 2 S52199  
280 25 100.0 347 2 S59837  
281 25 100.0 348 2 JN0633  
282 25 100.0 348 2 B81344  
283 25 100.0 350 2 C97346  
284 25 100.0 352 2 C86294  
285 25 100.0 353 2 S74379  
286 25 100.0 353 2 E81680  
287 25 100.0 353 2 A10670  
288 25 100.0 353 2 E64023  
289 25 100.0 356 2 A1888  
290 25 100.0 356 2 T37863  
291 25 100.0 358 2 D96579  
292 25 100.0 359 2 H95382  
293 25 100.0 360 1 S49188  
294 25 100.0 361 2 T45778  
295 25 100.0 362 2 S38637  
296 25 100.0 363 2 A31342  
297 25 100.0 363 2 G75050  
298 25 100.0 363 2 T20745  
299 25 100.0 363 2 B84932  
300 25 100.0 364 2 E71109

## ALIGNMENTS

RESULT 1  
S68260  
hypothetical protein gadd7.1 - long-tailed hamster  
C:Species: Cricetus longicaudatus (long-tailed hamster)  
C:Date: 06-Dec-1996 #sequence\_revision 13-Mar-1997 #text\_change 17-Mar-1999  
C:Accession: S68260  
R:Hollander, M.C.; Alamo, I.; Fornace Jr., A.J.  
Nucleic Acids Res. 24, 1589-1593, 1996  
A:Title: A novel DNA damage-inducible transcript, gadd7, inhibits cell growth, but lacks  
A:Reference number: S68260; MUID:96211359; PMID:8649973  
A:Accession: S68260  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-38 <HOL>  
A:Cross-references: EMBL:L40430

Query Match 100.0%; Score 25; DB 2; Length 38;  
Best Local Similarity 50.0%; Pred. NO. 70;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db 27 SEFITGDN 34  
:|||||:

## RESULT 2

E45681  
hypothetical 5.9K protein (gene 61.1 protein) - phase T4  
N:Alternate names: gp 61.1  
C:Species: phage T4  
A:Note: host *Escherichia coli*  
C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999  
C:Accession: E45681; J80562  
R:Selick, H.E.; Stormo, G.D.; Dyson, R.L.; Alberts, B.M.  
J. Virol. 67, 2305-2316, 1993  
A:Title: Analysis of five presumptive protein-coding sequences clustered between the pri  
A:Reference number: A45681; MUID:93188183; PMID:8383243  
A:Accession: E45681  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-54 <SEL>  
A:Cross-references: GB:S57514; NID:9298518; PIDN:AAB25712.1; PID:9298523  
A:Note: sequence extracted from NCBI backbone (NCBI:128289, NCBIP:128348)  
C:Genetics:  
A:Gene: 61.1  
A:Map position: 18.973-19.135

Query Match 100.0%; Score 25; DB 2; Length 54;  
Best Local Similarity 50.0%; Pred. NO. 1e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db 5 KEFIXEDM 12  
:|||||:

## RESULT 3

E81922  
hypothetical protein NMA0780 [imported] - *Neisseria meningitidis* (strain Z2491 serogroup  
C:Species: *Neisseria meningitidis*  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: E81922  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.  
A:Reference number: A81775; MUID:20222556; PMID:10761919  
A:Accession: E81922  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-67 <PAR>  
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84063.1; PID:g737950  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA0780

Query Match 100.0%; Score 25; DB 2; Length 67;  
Best Local Similarity 50.0%; Pred. NO. 1.3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db 16 YEFIRPDQ 23  
:|||||:

## RESULT 4

C96900  
small conserved protein, ortholog of YAAA B. subtilis CAC0003 [imported] - *Clostridium a*  
C:Species: *Clostridium acetobutylicum*

```
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: C96900
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C96900
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-68 <KUR>
A;Cross-references: GB:AF001437; PIDN:AAK77990.1; PID:g15022820; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0003

Query Match      100.0%; Score 25; DB 2; Length 68;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches      4; Conservative      4; Mismatches      0; Indels      0; Gaps      0;

Qy      1 XEFIXDX 8
      :|||:|:|:
Db      8 TEFIKLDS 15

RESULT 5
G97292
hypothetical protein CAC3193 [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: G97292
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: G97292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-79 <KUR>
A;Cross-references: GB:AF001437; PIDN:AAK81130.1; PID:g15026263; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3193

Query Match      100.0%; Score 25; DB 2; Length 79;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches      4; Conservative      4; Mismatches      0; Indels      0; Gaps      0;

Qy      1 XEFIXDX 8
      :|||:|:|:
Db      24 NEFINKDD 31

RESULT 6
WSWL13
E5 protein - human papillomavirus type 13
C;Species: human papillomavirus type 13
A;Note: nosc Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
C;Accession: F42955
R;van Ranst, M.; Fuse, A.; Fiten, P.; Beuken, E.; Pfister, H.; Burk, R.D.; Opdenakker, G
Virol. 190, 587-596, 1992
A;Title: Human papillomavirus type 13 and pygmy chimpanzee papillomavirus type 1: Compar
A;Reference number: A42955; MUID:92391075; PMID:1325697
A;Accession: F42955
A;Molecule type: DNA
A;Residues: 1-91 <VAN>
A;Cross-references: EMBL:X62843; NID:G60295; PIDN:CAA44652.1; PID:G60301
C;Superfamily: papillomavirus E5 protein
C;Keywords: early protein

Query Match      100.0%; Score 25; DB 1; Length 91;
```

```
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches      4; Conservative      4; Mismatches      0; Indels      0; Gaps      0;

Qy      1 XEFIXDX 8
      :|||:|:|:
Db      1 MEFIPIVDV 8

RESULT 7
S62338
L71-10 protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 24-Sep-1999
C;Accession: S62338; S62348
R;Wright, L.G.; Chen, T.; Thummel, C.S.; Guild, G.M.
J. Mol. Biol. 255, 387-400, 1996
A;Title: Molecular characterization of the 71E late puff in Drosophila melanogaster rev
A;Reference number: S62333; MUID:96152797; PMID:8568884
A;Accession: S62338
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-98 <WRI>
A;Cross-references: EMBL:U23836
A;Accession: S62348
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-98 <WRW>
A;Cross-references: EMBL:U24574; NID:g775244; PIDN:AAA65118.1; PID:g775245
C;Genetics:
A;Gene: L71-10
A;Cross-references: FlyBase:FBgn0014850
A;Introns: 78/1
C;Superfamily: L71-10 protein

Query Match      100.0%; Score 25; DB 2; Length 98;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches      4; Conservative      4; Mismatches      0; Indels      0; Gaps      0;

Qy      1 XEFIXDX 8
      :|||:|:|:
Db      46 LEFINSDC 53

RESULT 8
S12170
Hypothetical protein (3' terminal region) - carnation latent virus
C;Species: carnation latent virus, CLV
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 12-Jun-2003
C;Accession: S12170
R;Haylor, M.T.M.; Brunt, A.A.; Coutts, R.H.A.
Nucleic Acids Res. 18, 6127, 1990
A;Title: Conservation of the 3' terminal nucleotide sequence in five carlaviruses.
A;Reference number: S12170; MUID:91045079; PMID:2235495
A;Accession: S12170
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: genomic RNA
A;Residues: 1-101 <HAY>
A;Cross-references: EMBL:X55897; NID:g58887; PIDN:CAA39386.1; PID:g58888
C;Superfamily: potato virus nucleic acid-binding protein

Query Match      100.0%; Score 25; DB 2; Length 101;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches      4; Conservative      4; Mismatches      0; Indels      0; Gaps      0;

Qy      1 XEFIXDX 8
      :|||:|:|:
Db      89 LEFTIGIDL 96

RESULT 9
TI0321
hypothetical protein 52 - Orgyia pseudotsugata nuclear polyhedrosis virus
C;Species: Orgyia pseudotsugata nuclear polyhedrosis virus, OpMNPV
```

C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
 C:Accession: T10321  
 R:Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohrmann, G.F.  
 Virology 229, 381-399, 1997  
 A>Title: The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis  
 A:Reference number: Z17011; MUID:97271300; PMID:9126251  
 C:Accession: T10321  
 A>Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-109 <HR>  
 A:Cross-references: EMBL:U75930; NID:G2934903; PIDN:AAC59051.1; PID:G1911298  
 C:Superfamily: Autographa californica nuclear polyhedrosis virus EcoRI-T medium protein

Query Match 100.0%; Score 25; DB 2; Length 109;  
 Best Local Similarity 50.0%; Pred. No. 2.3e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||||:  
 Db 96 LEFIKLDV 103

RESULT 10  
 G85800  
 prophage pi3 protein 34 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
 C:Species: Lactococcus lactis subsp. lactis  
 C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: G85800  
 R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich  
 Genome Res. 11, 731-753, 2001  
 A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
 A:Reference number: A85625; MUID:21235186; PMID:11337471  
 C:Accession: G85800  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-114 <STO>  
 A:Cross-references: GB:AE005176; PID:G12724395; PIDN:AAK05505.1; GSPDB:GNO0146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: pi34

Query Match 100.0%; Score 25; DB 2; Length 114;  
 Best Local Similarity 50.0%; Pred. No. 2.5e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||||:  
 Db 22 GBFIVCDE 29

RESULT 11  
 A85636  
 hypothetical protein Z1420 [imported] - Escherichia coli (strain O157:H7, substrain EDL9  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: A85636  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 C:Accession: A85636  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-116 <STO>  
 A:Cross-references: GB:AE005174; NID:G12514272; PIDN:AAG55549.1; GSPDB:GNO0145; UWGP:Z14  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: Z1420

Query Match 100.0%; Score 25; DB 2; Length 116;  
 Best Local Similarity 50.0%; Pred. No. 2.5e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||||:  
 Db 11 PEFIFNDN 18

RESULT 12  
 Q08E8H  
 hypothetical 13K protein (transforming region) - human cytomegalovirus (strain AD169)  
 C:Species: human cytomegalovirus, human herpesvirus 5  
 C>Date: 20-Sep-1984 #sequence\_revision 20-Sep-1984 #text\_change 17-Jul-1998  
 C:Accession: A03800  
 R:Nelson, J.A.; Fleckenstein, B.; Jahn, G.; Galloway, D.A.; McDougall, J.K.  
 J. Virol. 49, 109-115, 1984  
 A>Title: Structure of the transforming region of human cytomegalovirus AD169.  
 A:Reference number: A03800; MUID:84090384; PMID:6317885  
 A:Accession: A03800  
 A:Molecule type: DNA  
 A:Residues: 1-118 <NEL>  
 C:Superfamily: human cytomegalovirus hypothetical 13K protein

Query Match 100.0%; Score 25; DB 1; Length 118;  
 Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||||:  
 Db 28 QEFITNDV 35

RESULT 13  
 F82462  
 conserved hypothetical protein VCA0423 [imported] - Vibrio cholerae (strain N16961 sero  
 C:Species: Vibrio cholerae  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, E  
 l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: A82035; MUID:20406833; PMID:10952301  
 C:Accession: F82462  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-119 <HEI>  
 A:Cross-references: GB:AE004376; GB:AE003853; NID:G9657811; PIDN:AAF96329.1; GSPDB:GNO01  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VCA0423  
 A:Map position: 2

Query Match 100.0%; Score 25; DB 2; Length 119;  
 Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||||:  
 Db 17 AEFIALDN 24

RESULT 14  
 D97300  
 probable membrane protein [imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C:Accession: D97300  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
 J. Bacteriol. 183, 4823-4838, 2001  
 A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 C:Accession: D97300

Db 93 VEFIVGDD 100

RESULT 17

JT0666

hypothetical 14.7K protein - African swine fever virus

N/Alternate names: open reading frame D129L

C/Species: African swine fever virus, ASFV

C/Date: 30-Sep-1993 #sequence\_revision 20-Aug-1994 #text\_change 21-Jul-2000

C/Accession: JT0666

R/Vanez, R.J.; Rodriguez, J.M.; Boursnell, M.; Rodriguez, J.F.; Vinuela, E.

Gene 134, 161-174, 1993

A/Title: Two putative African swine fever virus helicases similar to yeast 'DEAH' pre-mRNA

A/Reference number: JT0665; MUID:94085774; PMID:8262374

A/Accession: JT0666

A/Molecule type: DNA

A/Residues: 1-129 <YAN>

A/Cross-references: GB:U18466; NID:g780375; PIDN:AAA65332.1; PID:g780472

Query Match 100.0%; Score 25; DB 2; Length 129;

Best Local Similarity 50.0%; Pred.No. 2.8e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8

Db 120 TEFIKDD 127

RESULT 18

AH1093

hypothetical protein lmo0151 [imported] - Listeria monocytogenes (strain EGD-e)

C/Species: Listeria monocytogenes

C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C/Accession: AH1093

R/Glasz, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, A.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H. D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; May, O. C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.

A/Reference number: AB1077; MUID:21537279; PMID:11679669

A/Accession: AH1093

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-135 <GLA>

A/Cross-references: GB:NC\_003210; PIDN:CAC98366.1; PID:g16409510; GSPDB:GNC0177

A/Experimental source: strain EGD-e

C/Genetics:

A/Gene: lmo0151

Query Match 100.0%; Score 25; DB 2; Length 135;

Best Local Similarity 50.0%; Pred.No. 3e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8

Db 60 IEFIKDP 67

RESULT 19

T08492

probable polypeptide cyclase - Enterobacter aerogenes plasmid R751

C/Species: Enterobacter aerogenes

C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C/Accession: T08492

R/Thorsted, P.B.; Shah, D.S.; Macartney, D.; Kostelidou, K.; Thomas, C.M.

Plasmid 36, 95-111, 1996

A/Title: Conservation of the genetic switch between replication and transfer genes of I

A/Reference number: Z16434; MUID:97118926; PMID:8954881

A/Accession: T08492

A/Status: preliminary;

A/Molecule type: DNA

A/translated from GB/EMBL/DBD

A;Residues: 1-140 <THO>  
A;Cross-references: EMBL:U67194; NID:g1572520; PIDN:AAC64436.1.; PID:g1572541  
C;Genetics:  
A;Gene: tnpA  
A;Genome: plasmid R751

Query Match 100.0%; Score 25; DB 2; Length 140;  
Best Local Similarity 50.0%; Pred.No.3.le+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
DB 51 LEFIVDV 58  
:|||||:

RESULT 20  
G97212  
Probable acetyltransferase [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C;Accession: G97212  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cilo  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: G97212  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-149 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK80490.1; PID:g15025561; GSFPB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC2539

Query Match 100.0%; Score 25; DB 2; Length 149;  
Best Local Similarity 50.0%; Pred.No.3.3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
DB 24 LEFITGDS 31  
:|||||:

RESULT 21  
D81333  
Probable protein-tyrosine-phosphatase (EC 3.1.3.48) Cj1258 [imported] - Campylobacter je  
C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C;Accession: D81333  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrel  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A;Reference number: AB1250; MUID:20150912; PMID:10688204  
A;Accession: D81333  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-151 <PAR>  
A;Cross-references: GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CAB73512.1; PID:g696869  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: Cj1258  
C;Superfamily: protein-tyrosine-phosphatase, low molecular weight  
C;Keywords: phosphoric monoester hydrolase

Query Match 100.0%; Score 25; DB 2; Length 151;  
Best Local Similarity 50.0%; Pred.No.3.4e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
DB 18 AEFIMKDL 25  
:|||||:

A;Residues: 1-154 <STO>  
 A;Cross-references: GB:AE005673; NID:gl3422099; PIDN:AAK22829.1; GSPDB:GN00148  
 C;Genetics:  
 A;Gene: CC0844

Query Match 100.0%; Score 25; DB 2; Length 154;  
 Best Local Similarity 50.0%; Pred. No. 3.5e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 XEFIXDX 8  
 :|||:|:  
 Db 123 VEFIENDI 130

RESULT 25  
 B96920  
 uncharacterized conserved protein CAC0166 [imported] - Clostridium acetobutylicum  
 C;Species: Clostridium acetobutylicum  
 C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C;Accession: B96920  
 R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824  
 A;Reference number: A96900; PMID:21359325; PMID:21359325  
 A;Accession: B96920  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-155 <KUR>  
 A;Cross-references: GB:AE001437; PIDN:AAK78149.1; PID:gl5022995; GSPDB:GN00168  
 A;Experimental source: Clostridium acetobutylicum ATCC824  
 C;Genetics:  
 A;Gene: CAC0166

Query Match 100.0%; Score 25; DB 2; Length 155;  
 Best Local Similarity 50.0%; Pred. No. 3.5e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 XEFIXDX 8  
 :|||:|:  
 Db 19 KEFIFKDN 26

RESULT 26  
 S76232  
 hypothetical protein sl10272 - Synechocystis sp. (strain PCC 6803)  
 C;Species: Synechocystis sp.  
 A;Variety: PCC 6803  
 C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 C;Accession: S76232  
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 103-136, 1996  
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis s.  
 A;Reference number: S74322; PMID:97061201; PMID:8905231  
 A;Accession: S76232  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-157 <KAN>  
 A;Cross-references: EMBL:D90914; GB:AB001339; NID:gl653477; PIDN:BAAL8491.1; PID:gl65357  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C;Genetics:  
 A;Start codon: GTG  
 C;Superfamily: Synechocystis hypothetical protein sl10272

Query Match 100.0%; Score 25; DB 2; Length 157;  
 Best Local Similarity 50.0%; Pred. No. 3.5e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
 :|||:|:  
 Db 39 TEFTNQDF 46

## RESULT 27

AG1755  
 bacteriophage protein homolog lin2588 [imported] - Listeria innocua (strain Clp11262)  
 C;Species: Listeria innocua  
 C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
 C;Accession: AG1755  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Fsihi, H. Science 294, 849-852, 2001  
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M. ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland A;Title: Comparative genomics of Listeria species  
 A;Reference number: AB1077; PMID:21537279; PMID:11679669  
 A;Accession: AG1755  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-160 <GLA>  
 A;Cross-references: GB:AL592022; PIDN:CAC97815.1; PID:gl6415125; GSPDB:GN00178  
 A;Experimental source: strain Clp11262  
 C;Genetics:  
 A;Gene: lin2588  
 C;Superfamily: Streptococcus phage phi-O1205 hypothetical protein 11

Query Match 100.0%; Score 25; DB 2; Length 160;  
 Best Local Similarity 50.0%; Pred. No. 3.6e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
 :|||:|:  
 Db 35 AEFINIDL 42

## RESULT 28

C82083  
 hypothetical protein VC388 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
 C;Species: Vibrio cholerae  
 C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C;Accession: C82083  
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J. chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000  
 A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A;Reference number: AB2035; PMID:20408833; PMID:10952301  
 A;Accession: C82083  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-160 <HEI>  
 A;Cross-references: GB:AE004309; GB:AE003852; NID:g9656954; PIDN:AAF95531.1; GSPDB:GN00 A;Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C;Genetics:  
 A;Gene: VC2388  
 A;Map position: 1

Query Match 100.0%; Score 25; DB 2; Length 160;  
 Best Local Similarity 50.0%; Pred. No. 3.6e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
 :|||:|:  
 Db 26 REFTAQDR 33

## RESULT 29

D83883  
 hypothetical protein BH1868 [imported] - Bacillus halodurans (strain C-125)  
 C;Species: Bacillus halodurans  
 C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C;Accession: D83883  
 R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir

Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: D83883  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-163 <STO>  
A;Cross-references: GB:AF001513; GB:BA000004; NID:gi0174345; PIDN:BA05587.1; GSPDB:GN00  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH1868

Query Match 100.0%; Score 25; DB 2; Length 163;  
Best Local Similarity 50.0%; Pred. No. 3.7e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:|  
Db 11 REFIEDT 18

RESULT 30  
G81321  
Probable integral membrane protein Cj1165c [imported] - *Campylobacter jejuni* (strain NC1  
C;Species: *Campylobacter jejuni*  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Jun-2003  
C;Accession: G81321  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyp  
A;Reference number: A81250; MUID:20150912; PMID:10688204  
A;Accession: G81321  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-164 <PAR>  
A;Cross-references: GB:AL1139077; GB:AL111168; NID:g6968444; PIDN:CAB73419.1; PID:g696859  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: Cj1165c  
C;Superfamily: primosomal operon 14K protein

Query Match 100.0%; Score 25; DB 2; Length 164;  
Best Local Similarity 50.0%; Pred. No. 3.7e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:|  
Db 2 IEFILDM 9

RESULT 31  
G82910  
Peptide methionine sulfoxide reductase UU289 [imported] - *Ureaplasma urealyticum*  
C;Species: *Ureaplasma urealyticum*  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C;Accession: G82910  
R;Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
submitted to GenBank, February 2000  
A;Description: The complete sequence of *Ureaplasma urealyticum*: Alternate views of a min  
A;Reference number: A82870  
A;Accession: G82910  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-165 <GLA>  
A;Cross-references: GB:AE002125; GB:AF222894; NID:g6899253; PIDN:AAF30698.1; GSPDB:GN001  
A;Experimental source: serovar 3; biovar 1  
C;Genetics:  
A;Gene: msrA; UU289  
A;Genetic code: SGC3  
C;Superfamily: peptide methionine sulfoxide reductase

Query Match 100.0%; Score 25; DB 2; Length 165;

Best Local Similarity 50.0%; Pred. No. 3.7e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:|  
Db 105 DEFIIKDF 112

RESULT 32  
AE3385  
acetyltransferase (EC 2.3.1.1-) [imported] - *Brucella melitensis* (strain 16M)  
C;Species: *Brucella melitensis*  
C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C;Accession: AE3385  
R;DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
.; Mazur, M.; Goitsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A;Title: The genome sequence of the facultative intracellular pathogen *Brucella meliten*  
A;Reference number: AD3252; PMID:11756688  
A;Accession: AE3385  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-167 <KUR>  
A;Cross-references: GB:AE008917; PIDN:AAL52248.1; PID:gl7983034; GSPDB:GN00190  
A;Experimental source: strain 16M  
C;Genetics:  
A;Gene: BMEI1067  
A;Map position: 1  
C;Keywords: acyltransferase

Query Match 100.0%; Score 25; DB 2; Length 167;  
Best Local Similarity 50.0%; Pred. No. 3.8e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:|  
Db 57 SEFIVADN 64

RESULT 33  
G95987  
probable ureidoglycolate hydrolase (EC 3.5.3.19) [imported] - *Sinorhizobium meliloti* (st  
C;Species: *Sinorhizobium meliloti*  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 14-Sep-2001  
C;Accession: G95987  
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernar  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A;Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing end  
A;Reference number: A95842; MUID:21396508; PMID:11481431  
A;Accession: G95987  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-170 <KUR>  
A;Cross-references: GB:AL591985; PIDN:CAC49567.1; PID:gl5141054; GSPDB:GN00167  
A;Experimental source: strain 1021, megaplasmid pSymB  
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler,  
peta, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, B.; Komp, C.; Lelaure,  
hehault, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A;Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
A;Reference number: A96039; MUID:21368234; PMID:11474104  
C;Genetics:  
A;Gene: allA; Smb20873  
A;Genome: plasmid  
C;Keywords: hydrolase

Query Match 100.0%; Score 25; DB 2; Length 170;  
Best Local Similarity 50.0%; Pred. No. 3.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXXD 8  
:||||:|  
Db 136 SEFIWDR 143

## RESULT 34

G97106

uncharacterized conserved protein CAC1676 [imported] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C;Accession: G97106  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: G97106  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-170 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK79642.1; PID:gl5024638; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC1676

Query Match 100.0%; Score 25; DB 2; Length 170;  
Best Local Similarity 50.0%; Pred. No. 3.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXXD 8  
:||||:|  
Db 5 REFIGADV 12

## RESULT 35

E71530

hypothetical protein CT303 - Chlamydia trachomatis (serotype D, strain UW3/Cx)

C;Species: Chlamydia trachomatis  
C;Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C;Accession: E71530  
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998  
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis  
A;Reference number: A71570; MUID:99000809; PMID:9784136  
A;Accession: E71530  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-171 <ARN>  
A;Cross-references: GB:AE001303; GB:AE001273; NID:g3328718; PIDN:AAC67896.1; PID:g332871  
A;Experimental source: serotype D, strain UW-3/Cx  
C;Genetics:  
A;Gene: CT303

Query Match 100.0%; Score 25; DB 2; Length 171;  
Best Local Similarity 50.0%; Pred. No. 3.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXXD 8  
:||||:|  
Db 50 LEFIDNDY 57

## RESULT 36

C84141

hypothetical protein BH3931 [imported] - Bacillus halodurans (strain C-125)

C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: C84141  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: C84141

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-171 <STO>  
A;Cross-references: GB:AP001520; GB:BA000004; NID:gl0176401; PIDN:BA807650.1; GSPDB:GN00168  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH3931

Query Match 100.0%; Score 25; DB 2; Length 171;  
Best Local Similarity 50.0%; Pred. No. 3.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXXD 8  
:||||:|  
Db 137 NEFIWDS 144

## RESULT 37

CFYCBB

C-phycocyanin beta chain - Synechococcus sp.

C;Species: Synechococcus sp.  
C;Date: 13-Aug-1986 #sequence\_revision 13-Aug-1986 #text\_change 11-Jun-1999  
C;Accession: B94024; B94017; A00321; A22972  
R;de Lorimier, R.; Bryant, D.A.; Porter, R.D.; Liu, W.Y.; Jay, E.; Stevens Jr., S.E. Proc. Natl. Acad. Sci. U.S.A. 81, 7946-7950, 1984  
A;Title: Genes of the alpha and beta subunits of phycocyanin.  
A;Reference number: A94024; MUID:85088525; PMID:6096868  
A;Note: Agmenellum quadruplicatum

A;Accession: B94024  
A;Molecule type: DNA  
A;Residues: 1-172 <DEL>  
A;Cross-references: GB:K02650; NID:gl142182; PIDN:AAB05343.1; PID:gl142183  
A;Experimental source: strain PR-6  
R;Pilot, T.D.; Fox, J.L. Proc. Natl. Acad. Sci. U.S.A. 81, 6983-6987, 1984  
A;Title: Cloning and sequencing of the genes encoding the alpha and beta subunits of C-phycocyanin  
A;Reference number: A94017; MUID:85063716; PMID:6438628  
A;Note: Agmenellum quadruplicatum

A;Accession: B94017  
A;Molecule type: DNA  
A;Residues: 1-172 <PIL>  
A;Cross-references: GB:K02659; NID:gl142176; PIDN:AAB05341.1; PID:gl142177  
A;Experimental source: strain PR-6  
C;Superfamily: phycocyanin  
C;Keywords: chromoprotein; methylated amino acid; photosynthesis; phycocyanobilin  
F;72/Modified site: N4-methylasparagine (Asn) #status predicted  
F;82/Binding site: phycocyanobilin (Cys) (covalent) #status predicted  
F;153/Binding site: phycocyanobilin (Cys) (covalent) #status predicted

Query Match 100.0%; Score 25; DB 1; Length 172;  
Best Local Similarity 50.0%; Pred. No. 3.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXXD 8  
:||||:|  
Db 16 GEFISSDK 23

## RESULT 38

C82053

shikimate kinase VC2629 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C;Species: Vibrio cholerae  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C;Accession: C82053  
R;Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, I. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000  
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A;Reference number: A82035; MUID:20406833; PMID:10952301  
A;Accession: C82053  
A;Status: preliminary  
A;Molecule type: DNA



A;Residues: 1-174 <HEI>  
A;Cross-references: GB:AE004329; GB:AE003852; NID:G9657211; PIDN:AAF95770.1; GSPDB:GN00168  
A;Experimental source: serogroup O1; strain N16961; biotype EI Tor  
C;Genetics:  
A;Gene: VC2629  
A;Map position: 1  
C;Superfamily: shikimate kinase; shikimate kinase homology

Query Match 100.0%; Score 25; DB 2; Length 174;  
Best Local Similarity 50.0%; Pred. No. 4e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 4; Mismatches 0;

QY 1 XEFIXXDX 8  
:||||:|  
Db 30 MEFTDSDT 37

RESULT 39  
G90357  
dTDP-4-dehydrothamnose 3,5 epimerase (rfbC-2) [imported] - Sulfolobus solfataricus  
C;Species: Sulfolobus solfataricus  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001  
C;Accession: G90357  
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A;Description: Sulfolobus solfataricus complete genome.  
A;Reference number: A99139  
A;Accession: G90357  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-176 <KUR>  
A;Cross-references: GB:AE006641; NID:gl3815196; PIDN:AAK42118.1; GSPDB:GN00155  
C;Genetics:  
A;Gene: rfbC-2  
C;Superfamily: dTDP-4-dehydrothamnose 3,5-epimerase

Query Match 100.0%; Score 25; DB 2; Length 176;  
Best Local Similarity 50.0%; Pred. No. 4e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 4; Mismatches 0;

QY 1 XEFIXXDX 8  
:||||:|  
Db 155 KEFIISDK 162

RESULT 40  
D97084  
enzyme of dihydrofolate reductase family, ortholog YWFD B. subtilis [imported] - Clostridium  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C;Accession: D97084  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: D97084  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-176 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK79463.1; PID:gl5024442; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC1495

Query Match 100.0%; Score 25; DB 2; Length 176;  
Best Local Similarity 50.0%; Pred. No. 4e+02; Indels 0; Gaps 0;  
Matches 4; Conservative 4; Mismatches 0;

QY 1 XEFIXXDX 8  
:||||:|

Db 89 VEFINQDV 96

#### RESULT 41

H97205

probable kinase from adenylate kinase family, FLAR-like protein [imported] - Clostridium  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C;Accession: H97205  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: H97205  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-177 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK80435.1; PID:gl5025501; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC2481

#### Query Match

Best Local Similarity 100.0%; Score 25; DB 2; Length 177;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXXDX 8

:||||:|

Db 58 EEFIKMDK 65

#### RESULT 42

A64393

hypothetical protein MJ0745 - Methanococcus jannaschii

C;Species: Methanococcus jannaschii

C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000

C;Accession: A64393

R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;  
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.  
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A;Reference number: A64300; MUID:96337999; PMID:8688087  
A;Accession: A64393  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-181 <BUL>  
A;Cross-references: GB:U67520; GB:L77117; NID:gl591447; PIDN:AA898745.1; PID:gl591458; T  
C;Genetics:  
A;Map position: REV672219-671674  
A;Start codon: GTG  
C;Superfamily: hypothetical protein MJ0745

#### Query Match

Best Local Similarity 100.0%; Score 25; DB 2; Length 181;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXXDX 8

:||||:|

Db 23 SEFIITDD 30

#### RESULT 43

T48241

hypothetical protein T7H20.250 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C;Accession: T48241

R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; New  
submitted to the Protein Sequence Database, March 2000  
A;Reference number: Z24488

```
A;Accession: T48241
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-181 <BEV>
A;Cross-references: EMBL:AL162508
A;Experimental source: cultivar Columbia; BAC clone T7H20
C;Genetics:
A;Map position: 5
A;Intons: 158/1
A;Note: T7H20.250

Query Match      100.0%; Score 25; DB 2; Length 181;
Best Local Similarity 50.0%; Pred. No. 4.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 121 MEFIYHDS 128

RESULT 44
H72630
hypothetical protein APE1503 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: H72630
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; KDNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: H72630
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-181 <KAW>
A;Cross-references: DBJ:AP000061; NID:g5104821; PIDN:BAA80502.1; PID:d1044288; PID:g5104821
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1503

Query Match      100.0%; Score 25; DB 2; Length 181;
Best Local Similarity 50.0%; Pred. No. 4.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 100 IEFINNDP 107

RESULT 45
C83913
hypothetical protein BH2107 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: C83913
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; HiraNucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: C83913
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-183 <STO>
A;Cross-references: GB:AF001514; GB:BA000004; NID:g10174613; PIDN:BA805826.1; GSPDB:GN001514
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH2107

Query Match      100.0%; Score 25; DB 2; Length 183;
Best Local Similarity 50.0%; Pred. No. 4.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
```

```
Db 122 DERIQDE 129

RESULT 46
S78329
Photosystem I chain III - Odontella sinensis chloroplast
N;Alternate names: photosystem I plastocyanin-binding chain
C;Species: chloroplast Odontella sinensis
C;Date: 17-Feb-1998 #sequence_revision 26-Feb-1998 #text_change 20-Jun-2000
C;Accession: S78329
R;Kowalik, K.V.; Stoebe, B.; Schaffran, I.; Kroth-Pancic, P.; Freier, U.
Plant Mol. Biol. Rep. 13, 336-342, 1995
A;Title: The Chloroplast Genome of a chlorophyll a+c- containing Alga, Odontella sinensis
A;Reference number: S78328
A;Accession: S78329
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-185 <KOW>
A;Cross-references: EMBL:Z67753; NID:g1185127; PIDN:CAA91702.1; PID:g1185219
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995
C;Genetics:
A;Gene: psar
A;Genome: chloroplast
C;Superfamily: photosystem I chain III
C;Keywords: chloroplast; membrane-associated complex; photosynthesis; photosystem I

Query Match      100.0%; Score 25; DB 2; Length 185;
Best Local Similarity 50.0%; Pred. No. 4.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 166 QEFISNDL 173

RESULT 47
E69176
hypothetical protein MTH578 - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 08-Dec-1997 #text_change 22-Oct-1999
C;Accession: E69176
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiawani, N.; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: function
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: E69176
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-186 <MTH>
A;Cross-references: GB:AE000840; GB:AE000666; NID:g2621653; PIDN:AAB85084.1; PID:g2621653
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH578

Query Match      100.0%; Score 25; DB 2; Length 186;
Best Local Similarity 50.0%; Pred. No. 4.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 71 VEFISGDV 78

RESULT 48
G97069
precorrin-6B methylase Cbit [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 18-Aug-2003
C;Accession: G97069
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee
```

```

.: Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium
A:Accession: G97069
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-187 <KUR>
A/Cross-references: GB:AE001437; PIDN:AAK79346.1; PID:gl5024314; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
C:Superfamily: precorrin-6Y methylase CbiT

Query Match      100.0%; Score 25; DB 2; Length 187;
Best Local Similarity 50.0%; Pred. No. 4.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
:||||:
Db 7 EEFIRGDC 14

RESULT 49
C97705
hypothetical protein RC0043 [imported] - Rickettsia conorii (strain Malish 7)
C/Species: Rickettsia conorii
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C/Accession: C97705
R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
Science 293, 2093-2098, 2001
A>Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A/Reference number: A97700; MUID:21442074; PMID:11557893
A/Accession: C97705
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <KUR>
A/Cross-references: GB:AE006914; PIDN:AAL02581.1; PID:gl5619078; GSPDB:GN00173
C:Genetics:
C:Superfamily: hypothetical protein HI0304

Query Match      100.0%; Score 25; DB 2; Length 189;
Best Local Similarity 50.0%; Pred. No. 4.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
:||||:
Db 156 KEFIFADN 163

RESULT 50
A70100
hypothetical protein BB0001 - Lyme disease spirochete
C/Species: Borrelia burgdorferi (lyme disease spirochete)
C/Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 28-Jul-2000
C/Accession: A70100
R/Fraser, C.M.; Castjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A>Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A/Reference number: A70100; MUID:98065943; PMID:9403685
A/Accession: A70100
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-190 <KLE>
A/Cross-references: GB:AE001115; GB:AE000783; NID:g2687879; PIDN:AAC66406.1; PID:g268789
A:Experimental source: strain B31
C:Superfamily: Borrelia burgdorferi hypothetical protein BB0001

Query Match      100.0%; Score 25; DB 2; Length 190;

```

```

Best Local Similarity 50.0%; Pred. No. 4.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
:||||:
Db 177 KEFILLDL 184

RESULT 51
B47700
glycerol metabolism regulatory protein GlpP - Bacillus subtilis
C/Species: Bacillus subtilis
C/Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C/Accession: B47700; D69634
R/Beijer, L.; Nilsson, R.P.; Holmberg, C.; Rutberg, L.
J. Gen. Microbiol. 139, 349-359, 1993
A>Title: The glpP and glpF genes of the glycerol regulon in Bacillus subtilis.
A/Reference number: A47700; MUID:93171878; PMID:8436953
A/Accession: B47700
A:Molecule type: DNA
A:Residues: 1-192 <BEI>
A/Cross-references: GB:M99611; NID:gl42995; PIDN:AAA22489.1; PID:gl42996
A/Note: sequence extracted from NCBI backbone (NCBIN:125668, NCBI:P:125671)
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte:
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch:
A.; Ehrlich, S.D.; Emerson, P.F.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell:
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Serol
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, I
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: D69634
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-192 <KUN>
A/Cross-references: GB:Z99108; GB:Z99109; GB:AL009126; NID:g2633260; PIDN:CAB12766.1; PI
A:Experimental source: strain 168
C:Genetics:
C:Gene: glpP
A/Map position: 75 (degrees)
C/Keywords: transcription antitermination; transcription regulation

Query Match      100.0%; Score 25; DB 2; Length 192;
Best Local Similarity 50.0%; Pred. No. 4.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
:||||:
Db 71 AEPICQDI 78

RESULT 52
AD3014
conserved hypothetical protein Atu3720 [imported] - Agrobacterium tumefaciens (strain C58)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C/Accession: AD3014
R/Wood, D.W.; Stetubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193

```

A:Accession: AD3014  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-192 <KUR>  
A:Cross-references: GB:AB008689; PIDN:AAL44530.1; PID:g17742143; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu3720  
A:Map position: linear chromosome

Query Match 100.0%; Score 25; DB 2; Length 192;  
Best Local Similarity 50.0%; Pred. No. 4.4e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:  
Db 159 DEFIVGDD 166

RESULT 53  
B84829  
hypothetical protein At2g40440 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84829  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Unayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.W.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: B84829  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-194 <STO>  
A:Cross-references: GB:AE002093; NID:g2651312; PIDN:AAB87592.1; GSPDB:GN00139  
A:Gene: At2g40440  
A:Map position: 2

Query Match 100.0%; Score 25; DB 2; Length 194;  
Best Local Similarity 50.0%; Pred. No. 4.5e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:  
Db 86 IEFYSDG 93

RESULT 54  
S64583  
hypothetical protein YGR251w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein G9140  
C:Species: Saccharomyces cerevisiae  
C:Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 19-Apr-2002  
C:Accession: S64583  
R:Agostoni Carbone, M.L.; Panzeri, L.; Melchiorretto, P.; Carignani, G.; Feroli, F.; From  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64577  
A:Accession: S64583  
A:Molecule type: DNA  
A:Residues: 1-196 <AGO>  
A:Cross-references: EMBL:Z73036; NID:g1323455; PID:e243683; PID:g1323456; GSPDB:GN00007;  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: MIPS:YGR251w  
A:Cross-references: SGD:S0003483  
A:Map position: 7R

Query Match 100.0%; Score 25; DB 2; Length 196;  
Best Local Similarity 50.0%; Pred. No. 4.5e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:  
Db 93 GERIEGDX 100

## RESULT 55

D90669  
probable regulator [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050995  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: D90669  
R:Hayashi, T.; Makinaga, T.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
A:Reference number: A96629; MUID:21156231; PMID:11258796  
A:Accession: D90669  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-196 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA333747.1; PID:g13359781; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs0324

Query Match 100.0%; Score 25; DB 2; Length 196;  
Best Local Similarity 50.0%; Pred. No. 4.5e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:  
Db 72 TEFINRDN 79

## RESULT 56

G85519  
probable regulator ykgK [imported] - Escherichia coli (strain O157:H7, substrain EDL933  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: G85519  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhe  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dinalanta, E.; Potamouis, K.; Apodaca  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: G85519  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-196 <STO>  
A:Cross-references: GB:AF005174; NID:g12513080; PIDN:AA54619.1; GSPDB:GN00145; UMGF:Z070  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: ykgK

Query Match 100.0%; Score 25; DB 2; Length 196;  
Best Local Similarity 50.0%; Pred. No. 4.5e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:  
Db 72 TEFINRDN 79

## RESULT 57

F64755  
ykgK protein - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: F64755  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503  
 A;Accession: F64755  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-196 <BIAT>  
 A;Cross-references: GB:AB000136; GB:U00096; NID:g2367103; PIDN:AAC73397.1; PID:g2367106;  
 A;Experimental source: strain K-12, substrain MGL655  
 C;Genetics:  
 A;Gene: ykGK

Query Match 100.0%; Score 25; DB 2; Length 196;  
 Best Local Similarity 50.0%; Pred. No. 4.5e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
 :|||:|:  
 Db 72 SEFINRDN 79

RESULT 58  
 T02787  
 Probable submergence induced protein 2 - rice  
 C;Species: Oryza sativa (rice)  
 C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 24-Mar-1999  
 C;Accession: T02787  
 R;Lorbiecke, R.; Sauter, M.  
 submitted to the EMBL Data Library, February 1998  
 A;Reference number: Z14738  
 A;Accession: T02787  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-199 <LOR>  
 A;Cross-references: EMBL:AF050200; NID:g2952337; PID:g2952338  
 C;Genetics:  
 A;Gene: SIP2

Query Match 100.0%; Score 25; DB 2; Length 199;  
 Best Local Similarity 50.0%; Pred. No. 4.6e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
 :|||:|:  
 Db 34 KEFIPVDK 41

RESULT 59  
 S73365  
 Translation initiation factor IF-3 - Mycoplasma pneumoniae (strain ATCC 29342)  
 N;Alternate names: hypothetical protein C09\_orf201  
 C;Species: Mycoplasma pneumoniae  
 A;Variety: ATCC 29342  
 C;Date: 26-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999  
 C;Accession: S73365  
 R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkil, E.; Li, B.C.; Herrmann, R.  
 Nucleic Acids Res. 24, 4420-4449, 1996  
 A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
 A;Reference number: S73327; MUID:97105885; PMID:8948633  
 A;Accession: S73365  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-201 <HIM>  
 A;Cross-references: EMBL:AB000005; GB:U00089; NID:g1673684; PIDN:AB95687.1; PID:g167368  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
 C;Genetics:  
 A;Gene: infC  
 A;Genetic code: SGC3  
 C;Superfamily: translation initiation factor IF-3

Query Match 100.0%; Score 25; DB 2; Length 201;  
 Best Local Similarity 50.0%; Pred. No. 4.7e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8

Db 44 NEFILIDE 51  
 :|||:|:

RESULT 60  
 T50322  
 Hypothetical coiled-coil protein [imported] - fission yeast (Schizosaccharomyces pombe)  
 C;Species: Schizosaccharomyces pombe  
 C;Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
 C;Accession: T50322  
 R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Cadieu, E.; Lelaure, V.; Galibert,  
 submitted to the EMBL Data Library, January 2000  
 A;Reference number: Z25061  
 A;Accession: T50322  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-202 <MCD>  
 A;Cross-references: EMBL:AL136536; PIDN:CAB66453.1; GSPDB:GN00067; SPDB:SPBC1703.09  
 A;Experimental source: strain 972h(-); cosmid cl703  
 C;Genetics:  
 A;Gene: SPDB:SPBC1703.09  
 A;Map position: 2  
 A;Introns: 155/3

Query Match 100.0%; Score 25; DB 2; Length 202;  
 Best Local Similarity 50.0%; Pred. No. 4.7e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
 :|||:|:  
 Db 38 TEIPADV 45

RESULT 61  
 AB1527  
 Hypothetical protein lin0754 [imported] - Listeria innocua (strain Clip11262)  
 C;Species: Listeria innocua  
 C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C;Accession: AB1527  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Me  
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 A;Title: Comparative genomics of Listeria species.  
 A;Reference number: AB1077; MUID:21537279; PMID:11679669  
 A;Accession: AB1527  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-202 <GLA>  
 A;Cross-references: GB:AL592022; PIDN:CAC95986.1; PID:g16413205; GSPDB:GN00178  
 A;Experimental source: strain Clip11262  
 C;Genetics:  
 A;Gene: lin0754

Query Match 100.0%; Score 25; DB 2; Length 202;  
 Best Local Similarity 50.0%; Pred. No. 4.7e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
 :|||:|:  
 Db 33 AEFINSDA 40

RESULT 62  
 AH1169  
 Hypothetical protein lmo0760 [imported] - Listeria monocytogenes (strain EGD-e)  
 C;Species: Listeria monocytogenes  
 C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C;Accession: AH1169  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Me  
ok, C.; Schluerer, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AH1169  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-202 <GLA>  
A:Cross-references: GB:NC 003210; PIDN:CAC98838.1; PID:g16410149; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0760

Query Match 100.0%; Score 25; DB 2; Length 202;  
Best Local Similarity 50.0%; Pred. No. 4.7e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 33 AEFIAGDA 40

RESULT 63  
AC1595  
Hypothetical protein lin1300 [imported] - *Listeria innocua* (strain Clip11262)  
C:Species: *Listeria innocua*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AC1595  
R:Glaser, P.; Prangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluerer, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1595  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-203 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC96531.1; PID:g16413773; GSPDB:GN00178  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lin1300

Query Match 100.0%; Score 25; DB 2; Length 203;  
Best Local Similarity 50.0%; Pred. No. 4.7e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 57 NEFIVRDQ 64

RESULT 64  
F64353  
dCTP deaminase (EC 3.5.4.13) MJ0430 [similarity] - *Methanococcus jannaschii*  
C:Species: *Methanococcus jannaschii*  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: F64353  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C  
A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*  
A:Reference number: A64300; MUID:96337999; PMID:8688087  
A:Accession: F64353  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-204 <BUL>

A:Cross-references: GB:U67494; GB:L77117; NID:g1591120; PIDN:AAB98415.1; PID:g1591133;  
C:Genetics:  
A:Map position: RV386963-386349  
C:Superfamily: dCTP deaminase  
C:Keywords: hydrolase

Query Match 100.0%; Score 25; DB 2; Length 204;  
Best Local Similarity 50.0%; Pred. No. 4.8e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 38 DEFIYDD 45

RESULT 65  
G83060  
probable ribosomal protein L25 PA4671 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
C:Species: *Pseudomonas aeruginosa*  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Dec-2002  
C:Accession: G83060  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Lim  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic path  
A:Reference number: AB2950; MUID:20437337; PMID:10984043  
A:Accession: G83060  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-204 <STO>  
A:Cross-references: GB:AE004882; GB:AE004091; NID:g9950939; PIDN:AAG08058.1; GSPDB:GN00  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4671  
C:Superfamily: 50S ribosomal protein L25

Query Match 100.0%; Score 25; DB 2; Length 204;  
Best Local Similarity 50.0%; Pred. No. 4.8e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 144 PEFIEVDM 151

RESULT 66  
A28864  
protease B (EC 3.4.23.-) - fungus (*Scytalidium lignicolum*)  
C:Species: *Scytalidium lignicolum*  
C:Date: 28-Aug-1989 #sequence\_revision 28-Aug-1989 #text\_change 31-Dec-1993  
C:Accession: A28864  
R:Maita, T.; Nagata, S.; Matsuda, G.; Maruta, S.; Oda, K.; Murao, S.; Tsuru, D.  
J. Biochem. 95, 465-475, 1984  
A:Title: Complete amino acid sequence of *Scytalidium lignicolum* acid protease B.  
A:Reference number: A28864; MUID:84185536; PMID:6370989  
A:Accession: A28864  
A:Molecule type: protein  
A:Residues: 1-204 <MAI>  
C:Keywords: aspartic proteinase; hydrolase

Query Match 100.0%; Score 25; DB 2; Length 204;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 130 AEFIIEFD 137

RESULT 67  
C98270  
hypothetical protein AGR\_L\_2230 [imported] - *Agrobacterium tumefaciens* (strain C58, C

C;Species: Agrobacterium tumefaciens  
 C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
 C;Accession: C98270  
 R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 Science 294, 2323-2328, 2001  
 A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A;Reference number: A97359; MUID:21608551; PMID:11743194  
 A;Accession: C98270  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-205 <KUR>  
 A;Cross-references: GB:AE007870; PIDN:AAK9685.1; PID:gl5159591; GSPDB:GN00170  
 C;Genetics:  
 A;Gene: AGR L 2230  
 A;Map position: linear chromosome

Query Match 100.0%; Score 25; DB 2; Length 205;  
 Best Local Similarity 50.0%; Pred. No. 4.8e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||:|:  
 Db 172 DEFIGDD 179

RESULT 68  
 H71671  
 ribosomal protein L4 - Rickettsia prowazekii  
 C;Species: Rickettsia prowazekii  
 C;Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
 C;Accession: H71671  
 R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichert-Ponten, T.; Alsmark, U  
 Nature 396, 133-140, 1998  
 A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
 A;Reference number: A71630; MUID:99039499; PMID:9823893  
 A;Accession: H71671  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-207 <AND>  
 A;Cross-references: GB:AJ235272; GB:AJ235269; NID:G3861033; PIDN:CAA15098.1; PID:g386119  
 A;Experimental source: strain Madrid E  
 C;Genetics:  
 A;Gene: rplD; RP658  
 C;Superfamily: Escherichia coli ribosomal protein L4

Query Match 100.0%; Score 25; DB 2; Length 207;  
 Best Local Similarity 50.0%; Pred. No. 4.8e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||:|:  
 Db 25 VEFIRDDI 32

RESULT 69  
 E97825  
 50S ribosomal protein L4 [imported] - Rickettsia conorii (strain Malish 7)  
 C;Species: Rickettsia conorii  
 C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
 C;Accession: E97825  
 R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro  
 Science 293, 2093-2098, 2001  
 A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
 A;Reference number: A97700; MUID:21442074; PMID:11557893  
 A;Accession: E97825  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-207 <KUR>  
 A;Cross-references: GB:AE006914; PIDN:AAL03543.1; PID:gl5620120; GSPDB:GN00173  
 C;Genetics:  
 A;Gene: rplD  
 C;Superfamily: Escherichia coli ribosomal protein L4

Query Match 100.0%; Score 25; DB 2; Length 207;  
 Best Local Similarity 50.0%; Pred. No. 4.8e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||:|:  
 Db 25 VEFIRDDI 32

RESULT 70  
 F95892  
 probable glutathione S-transferase protein [imported] - Sinorhizobium meliloti (strain  
 C;Species: Sinorhizobium meliloti  
 C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 02-Aug-2002  
 C;Accession: F95892  
 R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herna  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
 A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing end  
 A;Reference number: A95842; MUID:21396508; PMID:11481431  
 A;Accession: F95892  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-208 <KUR>  
 A;Cross-references: GB:AL591985; PIDN:CAC48806.1; PID:gl5140279; GSPDB:GN00167  
 A;Experimental source: strain 1021, megaplasmid pSymB  
 R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler  
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 hebaute, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A;Reference number: A96039; MUID:21368234; PMID:11474104  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: Smb20420  
 A;Genome: plasmid  
 C;Superfamily: hypothetical protein b0838

Query Match 100.0%; Score 25; DB 2; Length 208;  
 Best Local Similarity 50.0%; Pred. No. 4.9e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||:|:  
 Db 148 REFIGDR 155

RESULT 71  
 D82532  
 50S ribosomal protein L25 XF2643 [imported] - Xylella fastidiosa (strain 9a5c)  
 C;Species: Xylella fastidiosa  
 C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Dec-2002  
 C;Accession: D82532  
 R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 Nature 406, 151-157, 2000  
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A;Reference number: A82515; MUID:20365717; PMID:10910347  
 A;Note: for a complete list of authors see reference number A59328 below  
 A;Accession: D82532  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-208 <SIM>  
 A;Cross-references: GB:AE004071; GB:AE003849; NID:gg107866; PIDN:AAF85440.1; GSPDB:GN001  
 A;Experimental source: strain 9a5c  
 R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carriaro, D.M.; Carrer, H  
 as-Neco, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.B.; Laigh  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.P.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshakho, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XF2643  
C;Superfamily: 50S ribosomal protein L25

Query Match 100.0%; Score 25; DB 2; Length 208;  
Best Local Similarity 50.0%; Pred. No. 4.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:  
Db 144 PEFIEVDL 151

RESULT 72  
TI7329  
hypotheical protein DKFP564K2364.1 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: TI7329  
R;Dueterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, September 1999  
A;Reference number: Z18727  
A;Accession: TI7329  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-208 <DUE>  
A;Cross-references: EMBL:AL117620  
A;Experimental source: fetal brain; clone DKFP564K2364  
C;Genetics:  
A;Note: DKFP564K2364.1

Query Match 100.0%; Score 25; DB 2; Length 208;  
Best Local Similarity 50.0%; Pred. No. 4.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:  
Db 159 VEFIRHDR 166

RESULT 73  
I40484  
uracil phosphoribosyltransferase (EC 2.4.2.9) upp - Bacillus subtilis  
C;Species: Bacillus subtilis  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C;Accession: I40484; A69729; S49364  
R;Martinussen, J.; Glaser, P.; Andersen, P.S.; Saxild, H.H.  
J. Bacteriol. 177, 271-274, 1995  
A;Title: Two genes encoding uracil phosphoribosyltransferase are present in Bacillus sub  
A;Reference number: I40473; MUID:95095982; PMID:7798145  
A;Accession: I40484  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-209 <RES>  
A;Cross-references: EMBL:Z38002; NID:9556877; PIDN:CAA86111.1; PID:9556887  
R;Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galle  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mause  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,  
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: A69729  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-209 <KUN>  
A;Cross-references: GB:Z99122; GB:AL009126; NID:g2636029; PIDN:CAB15706.1; PID:g2636214  
A;Experimental source: strain 168  
C;Genetics:  
A;Gene: upp  
C;Superfamily: uracil phosphoribosyltransferase upp  
C;Keywords: glycosyltransferase; pentosyltransferase

Query Match 100.0%; Score 25; DB 1; Length 209;  
Best Local Similarity 50.0%; Pred. No. 4.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:  
Db 125 REFIVDP 132

RESULT 74  
C71647  
hypotheical protein RP851 - Rickettsia prowazekii  
C;Species: Rickettsia prowazekii  
C;Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C;Accession: C71647  
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark,  
Nature 396, 133-140, 1998  
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A;Reference number: A71630; MUID:99039499; PMID:9823893  
A;Accession: C71647  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-209 <AND>  
A;Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861237; PIDN:CAA15275.1; PID:g386133  
A;Experimental source: strain Madrid E  
C;Genetics:  
A;Gene: RP851

Query Match 100.0%; Score 25; DB 2; Length 209;  
Best Local Similarity 50.0%; Pred. No. 4.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:  
Db 69 REFIDADY 76

RESULT 75  
H97864  
hypotheical protein RC1320 [imported] - Rickettsia conorii (strain Malish 7)  
C;Species: Rickettsia conorii  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001  
C;Accession: H97864  
R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ri  
Science 293, 2093-2098, 2001  
A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A;Reference number: A97700; MUID:21442074; PMID:11557893  
A;Accession: H97864  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-209 <KUR>  
A;Cross-references: GB:AE006914; PIDN:AAL03858.1; PID:g15620461; GSPDB:GN00173  
C;Genetics:  
A;Gene: RC1320

Query Match 100.0%; Score 25; DB 2; Length 209;  
Best Local Similarity 50.0%; Pred. No. 4.9e+02;



Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:|

Db 69 IEFIDADY 76

RESULT 76

E89009

Protein R08E5.1 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001

C;Accession: E89009

R;Anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A;Accession: E89009

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-209 <STO>

A;Cross-references: GB:chr\_V; PIDN:AAB52275.1; PID:G1938440; GSPDB:GN00023; CBSP:R08E5.1

C;Genetics:

A;Gene: R08E5.1

A;Map position: 5

Query Match 100.0%; Score 25; DB 2; Length 209;

Best Local Similarity 50.0%; Pred. No. 4.9e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:|

Db 157 LEFIQCDA 164

RESULT 77

T41553

thymidylate kinase - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000

C;Accession: T41553

R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.

submitted to the EMBL Data Library, June 1998

A;Reference number: Z22001

A;Accession: T41553

A;Status: preliminary; translated from GB/EMBL/DDRJ

A;Molecule type: DNA

A;Residues: 1-210 <WOO>

A;Cross-references: EMBL:AL023794; PIDN:CAA19357.1; GSPDB:GN00068; SPDB:SPCC70.07c

A;Experimental source: strain 972h-; cosmid c70

C;Genetics:

A;Gene: SPDB:SPCC70.07c

A;Map position: 3

A;Introns: 45/1; 110/3

C;Superfamily: dtmP kinase

Query Match 100.0%; Score 25; DB 2; Length 210;

Best Local Similarity 50.0%; Pred. No. 4.9e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:|

Db 176 LEFITLDA 183

RESULT 78

S28955

dtmP kinase (EC 2.7.4.9) - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 19-Jan-2001

C;Accession: S28955

R;Abaigar, L.T.; Yeh, Y.I.-S.; Jong, A.Y.

Biochim. Biophys. Acta 1132, 222-224, 1992

A;Title: Functional and structural conservation of Schizosaccharomyces pombe dtmP kinase

A;Reference number: S28955; MUID:93003330; PMID:13271149

A;Accession: S28955

A;Molecule type: mRNA

A;Residues: 1-210 <ABA>

A;Cross-references: EMBL:X65868; NID:G4973; PIDN:CAA46698.1; PID:G4974

C;Superfamily: dtmP kinase

C;Keywords: ATP; nucleotide binding; nucleotide biosynthesis; P-loop; phosphotransferas;

F;14-21/Region: nucleotide-binding motif A (P-loop)

Query Match 100.0%; Score 25; DB 2; Length 210;

Best Local Similarity 50.0%; Pred. No. 4.9e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:|

Db 175 LEFITLDA 182

RESULT 79

A85098

hypothetical protein AT4g09590 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001

C;Accession: A85098

R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring

Nature 402, 769-777, 1999

A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A;Reference number: A85001; MUID:20083488; PMID:10617198

A;Accession: A85098

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-211 <STO>

A;Cross-references: GB:NC\_001268; NID:G7267654; PIDN:CAB78082.1; GSPDB:GN00140

C;Genetics:

A;Gene: AT4g09590

A;Map position: 4

Query Match 100.0%; Score 25; DB 2; Length 211;

Best Local Similarity 50.0%; Pred. No. 4.9e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
:||||:|

Db 49 PEFILQDT 56

RESULT 80

C84775

probable harpin-induced protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C;Accession: C84775

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84775

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-211 <STO>

A;Cross-references: GB:AE002093; NID:G4510372; PIDN:AAD21460.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g35970

A;Map position: 2

Query Match 100.0%; Score 25; DB 2; Length 211;

Best Local Similarity 50.0%; Pred. No. 4.9e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 XEFIXDX 8
      :|||::|:
Db      49 PEFILQDT 56

RESULT 81
JC4883
scytalidopepsin B (EC 3.4.23.32) precursor - fungus (Scytalidium lignicolum) (fragment)
N:Alternate names: pepstatin-insensitive acid; scytalidium aspartic proteinase B
C:Species: Scytalidium lignicolum
C:Date: 10-Sep-1996 #sequence_revision 18-Oct-1996 #text_change 21-Jul-2000
C:Accession: JC4883; PC4182
B:Kakimori, T.; Yoshimoto, T.; Oyama, H.; Oda, N.; Gotoh, Y.; Oda, K.; Murao, S.; Tsuru,
  Biosci. Biotechnol. Biochem. 60, 1210-1211, 1996
A:Title: Nucleotide sequence of the gene encoding pepstatin-insensitive acid protease B,
A:Reference number: JC4883; MUID:96376187; PMID:8782420
A:Accession: JC4883
A:Molecule type: DNA
A:Residues: 1-211 <KAL>
A:Cross-references: DDBJ:D83963; NID:gl514972; PIDN:BAAL2157.1; PID:gl514973
A:Accession: PC4182
A:Molecule type: protein
A:Residues: 1-211 <KAZ>
C:Comment: This enzyme is insensitive to pepstatin.
C:Genetics:
A:Introns: 81/1
A:Keywords: aspartic proteinase; hydrolase
F:1-5/Domain: signal sequence (fragment) #status predicted <SIG>
F:6-211/Product: scytalidopepsin B #status predicted <MAT>

      Query Match      100.0%; Score 25; DB 2; Length 211;
      Best Local Similarity 50.0%; Pred. No. 4.9e+02;
      Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 XEFIXDX 8
      :|||::|:
Db      136 AEFITDF 143

RESULT 82
AB3521
conserved hypothetical protein PA0990 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: AB3521
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
  adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
  ; Lory, S.; Olson, M.V.
  Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: AB3521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-212 <STO>
A:Cross-references: GB:AE004532; GB:AE004091; NID:g9946896; PIDN:AG04379.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0990

      Query Match      100.0%; Score 25; DB 2; Length 212;
      Best Local Similarity 50.0%; Pred. No. 5e+02;
      Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 XEFIXDX 8
      :|||::|:
Db      45 AEFIASDQ 52

RESULT 83
EB2675
hypothetical protein XF1489 [imported] - Xylella fastidiosa (strain 9a5c)

```

---

```

C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: EB2675
R:anonymous, the Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ
  Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: EB2675
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <SIM>
A:Cross-references: GB:AE003978; GB:AE003849; NID:g9106510; PIDN:AAF84238.1; GSPDB:GN00
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;
  Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer,
  as-Neto, E.; Docena, C.; El-Dorzy, H.; Facincani, A.P.; Ferreira, A.J.S.
  submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh
  J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
  chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,
  A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
  , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, C.V.
  Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasa
  A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
  M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;
  A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1489

      Query Match      100.0%; Score 25; DB 2; Length 213;
      Best Local Similarity 50.0%; Pred. No. 5e+02;
      Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 XEFIXDX 8
      :|||::|:
Db      126 DEFINGDS 133

RESULT 84
A70198
general stress protein (etc) homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Dec-2002
C:Accession: A70198
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whit
  son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt
  ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
  Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: A70198
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-214 <KLE>
A:Cross-references: GB:AE001177; GB:AE000783; NID:g2688711; PIDN:AAC67123.1; PID:g26887
A:Experimental source: strain B31
C:Superfamily: 50S ribosomal protein L25

      Query Match      100.0%; Score 25; DB 2; Length 214;
      Best Local Similarity 50.0%; Pred. No. 5e+02;
      Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 XEFIXDX 8
      :|||::|:
Db      172 PEFIELDL 179

RESULT 85
EB9339
conserved hypothetical protein AF0717 - Archaeoglobus fulgidus

```

C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C:Accession: E69339  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: E69339  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-214 <KLE>  
 A:Cross-references: GB:AE001055; GB:AE000782; NID:g2689378; PIDN:AAB90523.1; PID:g264989

Query Match 100.0%; Score 25; DB 2; Length 214;  
 Best Local Similarity 50.0%; Pred. No. 5e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 Db 40 AEFIRADV 47

RESULT 86  
 T07595  
 Glutathione transferase (EC 2.5.1.18) homolog GSTL - potato  
 N:Alternate names: glutathione S-transferase; pathogenesis-related protein prpl-1  
 C:Species: Solanum tuberosum (potato)  
 C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 08-Oct-1999  
 A:Accession: T07595  
 R:Strittmatter, G.  
 submitted to the EMBL Data Library, May 1995  
 A:Reference number: Z16038  
 A:Accession: T07595  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-217 <STR>  
 A:Cross-references: EMBL:J03679; NID:g862344; PIDN:AAA68430.1; PID:g169549  
 A:Experimental source: cv. Datura  
 C:Genetics:  
 A:Gene: gstl  
 A:Introns: 103/3  
 A:Note: activated by fungal infection  
 C:Superfamily: auxin-induced protein  
 C:Keywords: transferase

Query Match 100.0%; Score 25; DB 2; Length 217;  
 Best Local Similarity 50.0%; Pred. No. 5.1e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 Db 29 YEFIEDL 36

RESULT 87  
 T01076  
 transcription factor TINY - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 22-Oct-1999  
 C:Accession: T01076  
 R:Huang, E.N.; de la Bastide, M.; Habermann, K.; Schutz, K.; Shekher, M.; Vil, M.D.; Pre  
 submitted to the EMBL Data Library, August 1998  
 A:Description: BAC F18A17 from chromosome V, containing TINY at 60.5 cm.  
 A:Reference number: Z14247  
 A:Accession: T01076  
 A:Status: translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-218 <HUA>  
 A:Cross-references: EMBL:AC005405; NID:g3406034; PID:g3406035

A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Gene: TINY  
 A:Map position: 5

Query Match 100.0%; Score 25; DB 2; Length 218;  
 Best Local Similarity 50.0%; Pred. No. 5.1e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 Db 177 NEFIFSDS 184

RESULT 88  
 D90179  
 hypothetical protein SSO361 [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
 C:Accession: D90179  
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.  
 arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
 submitted to GenBank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: D90179  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-218 <KUR>  
 A:Cross-references: GB:AE006641; NID:g13813505; PIDN:AAK40691.1; GSPDB:GNO0155  
 C:Genetics:  
 A:Gene: SSO361

Query Match 100.0%; Score 25; DB 2; Length 218;  
 Best Local Similarity 50.0%; Pred. No. 5.1e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 Db 85 REFIEGDE 92

RESULT 89  
 A69505  
 uridylylate kinase (pyrH) homolog - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Aug-2003  
 C:Accession: A69505  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: A69505  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-219 <KLE>  
 A:Cross-references: GB:AE000962; GB:AE000782; NID:g2689285; PIDN:AAB89213.1; PID:g264949  
 C:Superfamily: uridylylate kinase

Query Match 100.0%; Score 25; DB 1; Length 219;  
 Best Local Similarity 50.0%; Pred. No. 5.2e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 Db 124 AEFIRADV 131

A:Molecule type: DNA  
A:Residues: 1-222 <HLA>  
A:Cross-references: GB:BA000007; PIDN:BAB35697.1; PID:gl3361740; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: ECS2274

Query Match 100.0%; Score 25; DB 2; Length 222;  
Best Local Similarity 50.0%; Pred. No. 5.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 XEPIXDX 8  
Db :|||:|

Db 54 VERIQDQG 61

RESULT 93

AF1345  
hypothetical protein lmo2166 [imported] - *Listeria monocytogenes* (strain EGD-e)  
C:Species: *Listeria monocytogenes*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AF1345  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke  
r.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kraft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; M  
ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1345  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-223 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAD00244.1; PID:gl6411636; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo2166

Query Match 100.0%; Score 25; DB 2; Length 223;  
Best Local Similarity 50.0%; Pred. No. 5.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 XEPIXDX 8  
Db :|||:|

Db 171 XEILPDE 178

RESULT 94

WNV29  
AcOrf-16 DA26 protein - Autographa californica nuclear polyhedrosis virus  
N:Alternate names: orf1 protein  
C:Species: Autographa californica nuclear polyhedrosis virus, ACNVPV  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 24-Nov-1999  
C:Accession: A29891; A34765; A44221; H72851  
R:Guarino, L.A.; Summers, M.D.  
J. Virol. 62, 463-471, 1988  
A:Title: Functional mapping of Autographa californica nuclear polyhedrosis virus genes  
A:Reference number: A93041; MUID:88091055; PMID:2826808  
A:Accession: A29891  
A:Molecule type: DNA  
A:Residues: 1-225 <GUA>  
A:Cross-references: EMBL:M18857; NID:g332482; PIDN:AAA66808.1; PID:g808753  
R:O'Reilly, D.R.; Passarelli, A.L.; Goldman, I.F.; Miller, L.K.  
J. Gen. Virol. 71, 1029-1037, 1990  
A:Title: Characterization of the DA26 gene in a hypervariable region of the Autographa  
A:Reference number: A34765; MUID:90264832; PMID:2189022  
A:Accession: A34765  
A:Molecule type: DNA  
A:Residues: 1-225 <OR>  
A:Cross-references: GB:M22619; GB:M33904; NID:g332422; PIDN:AAA69846.1; PID:g896308  
R:Braunagel, S.C.; Daniel, K.D.; Reilly, L.M.; Guarino, L.A.; Hong, T.; Summers, M.D.  
Virology 191, 1003-1008, 1992

A;Title: Sequence, genomic organization of the EcoRI-A fragment of Autographa californica VP8 of rotavirus.  
A;Reference number: A44221; MUID:93079853; PMID:1333113  
A;Accession: A44221  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-225 <BRA>  
A;Cross-references: GB:S52569  
R;Ayres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.  
Virology 202, 586-605, 1994  
A;Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.  
A;Reference number: A72850; MUID:94303173; PMID:8030224  
A;Accession: H72851  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-225 <AVR>  
A;Cross-references: GB:I22858; NID:G510708; PIDN:AAA66646.1; PID:G559085  
C;Comment: The gene is located in a 3-kilobase XbaI genome fragment.  
C;Comment: This protein is required for viral late gene expression.  
C;Genetics:  
A;Gene: ACOF-16  
C;Superfamily: Autographa californica nuclear polyhedrosis virus DA26 protein  
C;Keywords: early protein

Query Match 100.0%; Score 25; DB 1; Length 225;  
Best Local Similarity 50.0%; Pred. No. 5.3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 195 FEFINPDE 202

RESULT 95  
T23611  
hypothetical protein K11E4.2 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C;Accession: T23611  
R;Gajadaty, S.  
submitted to the EMBL Data Library, March 1996  
A;Reference number: Z19770  
A;Accession: T23611  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-226 <N1L>  
A;Cross-references: EMBL:Z70211; PIDN:CAA94158.1; GSPDB:GN00028; CESP:K11E4.2  
A;Experimental source: clone K11E4  
C;Genetics:  
A;Gene: CESP:K11E4.2  
A;Map position: X  
A;Introns: 30/3; 81/3; 137/3; 175/2

Query Match 100.0%; Score 25; DB 2; Length 226;  
Best Local Similarity 50.0%; Pred. No. 5.3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 170 GEFIIKRS 177

RESULT 96  
AB1477  
hypothetical protein lin0353 [imported] - Listeria innocua (strain Clp11262)  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AB1477  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, C.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland.  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AB1477  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-226 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC95586.1; PID:gl6412782; GSPDB:GN00178  
A;Experimental source: strain Clp11262  
C;Genetics:  
A;Gene: lin0353

Query Match 100.0%; Score 25; DB 2; Length 226;  
Best Local Similarity 50.0%; Pred. No. 5.3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 106 IEFIGEDT 113

RESULT 97  
AI1115  
hypothetical protein lmo0328 [imported] - Listeria monocytogenes (strain EGD-e)  
C;Species: Listeria monocytogenes  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AI1115  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AI1115  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-226 <GLA>  
A;Cross-references: GB:NC\_003210; PIDN:CAD00855.1; PID:gl6409692; GSPDB:GN00177  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: lmo0328

Query Match 100.0%; Score 25; DB 2; Length 226;  
Best Local Similarity 50.0%; Pred. No. 5.3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 106 IEFIGEDT 113

RESULT 98  
D90057  
hypothetical protein SA2319 [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: D90057  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A;Reference number: AB9758; MUID:21311952; PMID:11418146  
A;Accession: D90057  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-226 <KUR>  
A;Cross-references: GB:BA000018; PID:gl3702481; PIDN:BA843622.1; GSPDB:GN00149  
A;Experimental source: strain N315  
C;Genetics:  
A;Gene: SA2319

Query Match 100.0%; Score 25; DB 2; Length 226;  
Best Local Similarity 50.0%; Pred. No. 5.3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 185 DEFIKIDE 192

## RESULT 99

C71246  
hypothetical protein PH0226 - Pyrococcus horikoshii  
C;Species: Pyrococcus horikoshii  
C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 21-Jul-2000  
C;Accession: C71246  
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
A;Reference number: A71000; MUID:98344137; PMID:9679194  
A;Accession: C71246  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-227 <XAW>  
A;Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29298.1; PID:g3256615  
A;Experimental source: strain Or3  
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C;Genetics:  
A;Gene: PH0226  
C;Superfamily: conserved hypothetical protein MU1123; bioC homology  
P;39-144/Domain: bioC homology <BIOC>

Query Match 100.0%; Score 25; DB 2; Length 227;  
Best Local Similarity 50.0%; Pred. No. 5.4e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 87 VEFIVGDA 94

## RESULT 100

C75210  
sterol biosynthesis methyltransferase related PAB2213 - Pyrococcus abyssi (strain Orsay)  
C;Species: Pyrococcus abyssi  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C;Accession: C75210  
R;anonymus, Genoscope  
submitted to the EMBL Data Library, July 1999  
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru  
A;Reference number: A75001  
A;Accession: C75210  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-227 <XAW>  
A;Cross-references: GB:A248283; GB:AL096836; NID:g5457433; PIDN:CAB49130.1; PID:g545763  
A;Experimental source: strain Orsay  
C;Genetics:  
A;Gene: PAB2213  
C;Superfamily: spore germination protein C2; bioC homology

Query Match 100.0%; Score 25; DB 2; Length 227;  
Best Local Similarity 50.0%; Pred. No. 5.4e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 87 VEFIIIGDA 94

Search completed: September 12, 2004, 02:39:01  
Job time : 59 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2004, 21:30:41 ; Search time 44 Seconds  
(without alignments)

9.467 Million cell updates/sec

Title: US-09-660-302E-1

Perfect score: 25

Sequence: 1 XEFIXDX 8

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	25	100.0	54	1 Y01K_BPT4	P39228 bacterioph
2	25	100.0	91	1 VES_HPV13	Q02267 human papil
3	25	100.0	95	1 MGBB_HUMAN	Q75556 homo sapien
4	25	100.0	101	1 VNBP_CLV	P26625 carnation 1
5	25	100.0	109	1 VETM_NVP0P	Q10307 orgyia pseu
6	25	100.0	118	1 Y13K_HCMVA	P03336 human cytom
7	25	100.0	129	1 NUSE_STAEP	Q8cp37 staphylococ
8	25	100.0	146	1 STM2_MOUSE	P83093 mus musculus
9	25	100.0	151	1 Y4JR_RHISN	P55518 rhizobium s
10	25	100.0	165	1 MSRA_UREPA	Q9pgk2 ureaplasma
11	25	100.0	170	1 ALL1_RHIME	Q92ug4 rhizobium m
12	25	100.0	172	1 PHCB_SYNP2	P03944 rhizobium s
13	25	100.0	174	1 AROK_VIBCH	Q9kxv1 vibrio chol
14	25	100.0	175	1 Y05D_BPT4	P39259 bacterioph
15	25	100.0	185	1 PSFA_ODOSI	P49483 odontella s
16	25	100.0	190	1 Y001_BORBU	O51035 borellia bu
17	25	100.0	192	1 GLPP_BACSU	P30300 bacillus su
18	25	100.0	192	1 Y4DX_RHISN	P55423 rhizobium s
19	25	100.0	196	1 YG5C_YEAST	P53317 saccharomyc
20	25	100.0	196	1 YGKQ_ECOLI	P71301 escherichia
21	25	100.0	201	1 IF3_MYCPN	P78024 mycoplasma
22	25	100.0	204	1 DCD_METJA	Q57872 m bifunctio
23	25	100.0	206	1 RL25_THETH	P56930 thermus the
24	25	100.0	207	1 RL4_RICPR	Q92c96 rickettsia
25	25	100.0	209	1 UPP_BACSU	P39149 bacillus su
26	25	100.0	210	1 KTHY_SCHPO	P36590 schizosacch
27	25	100.0	213	1 URE1_HELMU	P50044 helicobacte
28	25	100.0	217	1 GTX1_SOLTU	P32111 solanum tub
29	25	100.0	219	1 PYRH_ARCFU	O28327 archaeoglob
30	25	100.0	225	1 VE26_NPVAC	P12827 autographa
31	25	100.0	226	1 PYRH_METJA	Q58656 methanococ
32	25	100.0	227	1 ORN_ARATH	Q9zve0 arabidopsis
33	25	100.0	228	1 B1OW_STAEP	Q8ctz3 staphylococ

34	25	100.0	229	1 RPIA_PYEFU	Q8ulf0 pyrococcus
35	25	100.0	229	1 UBIE_THEMA	Q9wz12 thermotoga
36	25	100.0	231	1 YF07_MYCTU	P71786 mycobacteri
37	25	100.0	233	1 RPIA_CHLCV	Q822p7 chlamydomph
38	25	100.0	236	1 VHEL_WCMVM	P09499 white clove
39	25	100.0	236	1 VHEL_WCMVM	P15403 white clove
40	25	100.0	236	1 YM08_SULSO	Q97w48 sulfolobus
41	25	100.0	240	1 YNM9_YEAST	P53915 saccharomyc
42	25	100.0	250	1 COX2_NEUCR	P00411 neurospora
43	25	100.0	250	1 GPP2_YEAST	P40106 saccharomyc
44	25	100.0	250	1 RS3_ALDYE	Q8vs63 alder yello
45	25	100.0	250	1 RS3_ELMPH	Q8vle1 elm witches
46	25	100.0	250	1 RS3_ELADO	Q8vle2 flavescence
47	25	100.0	250	1 RS3_RUBST	Q8vs60 rubus stunt
48	25	100.0	250	1 RS3_ZIZJU	Q8vs61 ziziphus ju
49	25	100.0	252	1 COX2_EMENI	P13588 emericella
50	25	100.0	256	1 Y0XB_BACSU	P28671 bacillus su
51	25	100.0	257	1 DPHB_SULSO	Q97tx8 sulfolobus
52	25	100.0	260	1 ARG7_SALTY	P02911 salmonella
53	25	100.0	260	1 PRTE_SCYLI	P15369 scytalidium
54	25	100.0	262	1 PHUF_ECOLI	P39405 escherichia
55	25	100.0	262	1 RNC_ARATH	P49204 arabidopsis
56	25	100.0	263	1 RNC_BACHD	Q9ka05 bacillus ha
57	25	100.0	269	1 MIND_ECOLI	P18197 escherichia
58	25	100.0	270	1 IL1A_RAT	P16598 rattus norv
59	25	100.0	271	1 IL1A_HUMAN	P1583 homo sapien
60	25	100.0	271	1 IL1A_MACFA	P79340 macaca fasc
61	25	100.0	271	1 IL1A_MACFA	P48089 macaca mula
62	25	100.0	272	1 TYRO_STRLN	P55023 streptomyce
63	25	100.0	275	1 NADE_STAEP	O8cnp1 staphylococ
64	25	100.0	275	1 TYRO_STRAL	P55022 streptomyce
65	25	100.0	277	1 NADE_CORGL	Q8mnn7 corynebacte
66	25	100.0	278	1 KC2B_YEAST	P43639 saccharomyc
67	25	100.0	280	1 TRMB_BACTN	Q8a0x7 bacteroides
68	25	100.0	281	1 YCX1_EUGGR	P31918 euglena gra
69	25	100.0	282	1 END4_WOLSU	Q7mb4 wolinnella s
70	25	100.0	282	1 Y079_RICCN	Q82j18 rickettsia
71	25	100.0	283	1 SPEE_CLOPE	Q8xny8 clostridium
72	25	100.0	284	1 CELA_ACEXY	P27897 acetobacter
73	25	100.0	288	1 RFBA_NEIMA	P57040 neisseria m
74	25	100.0	288	1 RFBA_NEIMB	P55255 neisseria m
75	25	100.0	289	1 RMLA_STRUM	P95778 streptococ
76	25	100.0	290	1 ARV2_MOUSE	P50295 mus musculu
77	25	100.0	291	1 ENGC_STAAC	Q9kx08 staphylococ
78	25	100.0	291	1 ENGC_STAAC	Q99up7 staphylococ
79	25	100.0	292	1 RFBA_SALTY	P55254 salmonella
80	25	100.0	292	1 RFBA_SALTY	P26393 salmonella
81	25	100.0	292	1 RFBA_SHIFL	P37779 shigella fl
82	25	100.0	293	1 RBAL_ECOLI	P37744 escherichia
83	25	100.0	296	1 FTR_METKA	Q49610 methanopyru
84	25	100.0	297	1 HSLO_CLOAB	Q97g16 clostridium
85	25	100.0	298	1 VPO_BHPH1	P51719 bacterioph
86	25	100.0	298	1 YL44_AQUAE	O67900 aquifex aeo
87	25	100.0	305	1 Y41F_RHISN	P55489 rhizobium s
88	25	100.0	306	1 COAA_LACLA	Q9c6m3 lactococcus
89	25	100.0	306	1 COAA_STA3	Q8es32 streptococ
90	25	100.0	308	1 OPPF_BACSU	P24137 bacillus su
91	25	100.0	309	1 MDH_CHLAU	P80040 chloroflexu
92	25	100.0	316	1 ANXC_HYDAT	P26256 hydra atten
93	25	100.0	317	1 RPOA_AQUAE	O66483 aquifex aeo
94	25	100.0	321	1 CVF_FSTNU	Q8w107 psilotum nu
95	25	100.0	321	1 FCB2_HUMAN	P06734 homo sapien
96	25	100.0	323	1 UME3_YEAST	P47821 saccharomyc
97	25	100.0	326	1 TF2B_ARCFU	Q28970 archaeoglob
98	25	100.0	327	1 RIR2_ASFM2	P26713 african swi
99	25	100.0	328	1 ASCD_YERPE	P37911 yersinia pe
100	25	100.0	329	1 LEU3_YETTH	O27441 methanobact
101	25	100.0	333	1 KC2B_NEUCR	Q8c912 neurospora
102	25	100.0	334	1 RIR2_ASFB7	P42492 african swi
103	25	100.0	336	1 F16P_SHEEP	P09199 ovine aries
104	25	100.0	336	1 GSPK_VIBCH	P45781 vibrio chol
105	25	100.0	337	1 F16P_HUMAN	P09467 homo sapien
106	25	100.0	337	1 F16P_MOUSE	Q9qxd6 mus musculu

107	25	100.0	337	1	FL6P_PIG	P00636 sus scrofa	180	25	100.0	431	1	SOT1_YEAST	P35184 saccharomyc
108	25	100.0	337	1	FL6P_RABIT	P00637 oryctolagus	181	25	100.0	433	1	ENGA_AQUAE	O67749 aquifex aeo
109	25	100.0	340	1	RFC2_SCHPO	Q09843 schizosacch	182	25	100.0	434	1	GID_FUSNN	Q8ren9 fusobacteri
110	25	100.0	341	1	SUR6_XENIA	Q01850 xenopus lae	183	25	100.0	439	1	YSCN_YERN	P40290 yersinia en
111	25	100.0	342	1	ARIS_PENRO	Q03471 penicillium	184	25	100.0	439	1	FLII_BACSU	P240291 bacillus ps
112	25	100.0	342	1	HEMZ_RICPR	Q02c84 rickettsia	185	25	100.0	440	1	FTSA_ENTFA	P23445 bacillus su
113	25	100.0	345	1	DDL_CAMJE	Q09pc2 campylobact	186	25	100.0	440	1	RUMA_XLFTA	O07111 enterococcu
114	25	100.0	347	1	RH31_YEAST	Q06624 saccharomyc	187	25	100.0	443	1	RUMA_XLFTA	O9pav7 xylella fas
115	25	100.0	347	1	PL11_WTV	P13094 wound tumor	188	25	100.0	443	1	RUMA_XLFTA	O87b33 xylella fas
116	25	100.0	348	1	PAP3_CARPA	P10056 carica papa	189	25	100.0	444	1	RUMA_XANCP	O8pmu6 xanthomonas
117	25	100.0	351	1	DDL_PROMA	Q7vas4 prochloroco	190	25	100.0	444	1	RUMA_XANCP	Q8pmu6 xanthomonas
118	25	100.0	353	1	MRP_SYNY3	P53383 synechocyst	191	25	100.0	446	1	AK_RICPR	Q9zci7 rickettsia
119	25	100.0	353	1	V634_CHLMU	Q9pk39 chlamydia m	192	25	100.0	446	1	BGLA_CLOTM	Q64163 mus musculu
120	25	100.0	353	1	YCS4_HABIN	P44140 haemophilus	193	25	100.0	448	1	BGLA_CLOTM	P26208 Clostridium
121	25	100.0	353	1	YNCE_SALTI	Q8z740 salmonella	194	25	100.0	448	1	VDR_COTUA	Q93611 homo sapien
122	25	100.0	353	1	YNCE_SALTI	Q8zpd6 salmonella	195	25	100.0	448	1	VDR_COTUA	Q93611 homo sapien
123	25	100.0	356	1	YD6B_SCHPO	Q10323 schizosacch	196	25	100.0	451	1	SFS2_MOUSE	O42392 gallus gall
124	25	100.0	358	1	TRMU_RICPR	Q9zdm1 rickettsia	197	25	100.0	452	1	SPS2_MOUSE	P97364 mus musculu
125	25	100.0	360	1	FIXB_AZOVI	P51574 azotobacter	198	25	100.0	453	1	VRK3_MOUSE	Q8k395 mus musculu
126	25	100.0	362	1	FL6P_RAT	P19112 rattus norv	199	25	100.0	454	1	MTW3_MICVA	P14244 micrococcus
127	25	100.0	365	1	Y550_METJA	Q57970 methanococc	200	25	100.0	455	1	HYIN_PSESS	P06618 pseudomonas
128	25	100.0	366	1	DE04_LACPL	Q8sv07 lactobacill	201	25	100.0	457	1	MMB1_METBA	O30642 methanosarc
129	25	100.0	370	1	TRMU_RICCN	Q92110 rickettsia	202	25	100.0	457	1	MMB2_METBA	Q9p914 methanosarc
130	25	100.0	372	1	BIOF_METJA	Q58694 methanococc	203	25	100.0	459	1	TIG_MYCPU	Q98r01 mycoplasma
131	25	100.0	375	1	GLK1_TRIVA	Q9gtw9 trichomonas	204	25	100.0	460	1	ACDB_METTH	O27745 methanobact
132	25	100.0	375	1	MCK1_YEAST	P21965 saccharomyc	205	25	100.0	465	1	TRME_CYACA	O9t1x6 cyanidium c
133	25	100.0	375	1	NPL4_HUMAN	Q99733 homo sapien	206	25	100.0	467	1	EUTA_SALTY	O9zfv2 salmoneilla
134	25	100.0	376	1	METK_AQUAE	O67222 aquifex aeo	207	25	100.0	468	1	COBQ_THEAC	Q9hl28 thermoplasm
135	25	100.0	377	1	GLK2_TRIVA	Q9gtw8 trichomonas	208	25	100.0	472	1	YCKJ_ECOLI	P78061 escherichia
136	25	100.0	378	1	DCAM_SCHPO	Q9p7e3 schizosacch	209	25	100.0	474	1	VRK3_HUMAN	O91613 homo sapien
137	25	100.0	380	1	PHLC_TRYCR	O15886 trypanosoma	210	25	100.0	477	1	MYVP_MOUSE	O45666 caenorhabdi
138	25	100.0	381	1	Y438_METJA	Q57880 methanococc	211	25	100.0	483	1	QUIC_ACICA	O43922 acinetobact
139	25	100.0	382	1	ISCL_ARCFU	O30052 archaeoglob	212	25	100.0	486	1	ENGA_SHEON	Q8ec36 shewanella
140	25	100.0	385	1	KUPL_CAEEL	P34705 caenorhabdi	213	25	100.0	487	1	ENGA_SHEON	P35005 zea mays (m
141	25	100.0	392	1	SESI_HUMAN	P49903 homo sapien	214	25	100.0	488	1	AMYB_MAIZE	Q7ff59 escherichia
142	25	100.0	392	1	SPS1_MOUSE	Q8bh69 mus musculu	215	25	100.0	490	1	ENGA_ECOLI	P77254 escherichia
143	25	100.0	392	1	THIL_CHRVO	Q9zhi1 chromobacte	216	25	100.0	490	1	ENGA_SALTY	O8z4p6 salmoneilla
144	25	100.0	393	1	PURT_METJA	Q58881 methanococc	217	25	100.0	490	1	ENGA_SALTY	O9xcia salmoneilla
145	25	100.0	399	1	ZOT_VIBCH	P38442 vibrio chol	218	25	100.0	492	1	GRXA_BACAN	O9zfb4 bacillus an
146	25	100.0	401	1	CSD_PYRAB	Q9v242 pyrococcus	219	25	100.0	492	1	Y100_STEPR	Q7p744 streptococc
147	25	100.0	402	1	ROO_DESGI	Q9f0j6 desulfovibr	220	25	100.0	493	1	Y100_STEPR	Q9ktw7 vibrio chol
148	25	100.0	404	1	METK_CHLTE	Q8k67 chlorobium	221	25	100.0	494	1	ENGA_YERPE	Q8zct9 yersinia pe
149	25	100.0	405	1	DAPE_STAPP	Q8qcc2 staphylococ	222	25	100.0	495	1	ENGA_YERPE	O67298 aquifex aeo
150	25	100.0	405	1	S2AE_BACSU	P49782 bacillus su	223	25	100.0	497	1	SYM_AQUAE	O87s12 vibrio para
151	25	100.0	406	1	CLPX_THEMEA	Q9wx23 thermotoga	224	25	100.0	498	1	ENGA_VIBPA	Q98127 schizosacch
152	25	100.0	406	1	YAS9_METJA	Q58459 methanococc	225	25	100.0	498	1	YAC3_SCHPO	Q91tm0 arabidopsis
153	25	100.0	412	1	CLPX_AQUAE	O67356 aquifex aeo	226	25	100.0	501	1	AMYB_WHEAT	P93594 triticum ae
154	25	100.0	412	1	FNG_DROME	Q24342 drosophila	227	25	100.0	503	1	YF26_SYNY3	P74360 synechocyst
155	25	100.0	413	1	APGM_SULTO	Q975p3 sulfolobus	228	25	100.0	506	1	CBP3_HORVU	P21529 hordeum vul
156	25	100.0	414	1	CLPX_NEIMA	O91tx8 neisseria m	229	25	100.0	508	1	GP2_CANFA	P25291 canis fami
157	25	100.0	414	1	CLPX_NEIME	Q91yy3 neisseria m	230	25	100.0	509	1	GP2_CANFA	O8tmb3 lilium cana
158	25	100.0	417	1	CLPX_CAMJE	Q9pim0 campylobact	231	25	100.0	512	1	MATK_LILCA	O9gib3 lilium henr
159	25	100.0	418	1	YCCF_ECOLI	P36661 escherichia	232	25	100.0	512	1	MATK_LILRE	O9ghc3 lilium rega
160	25	100.0	419	1	DNLI_ASEF7	P25970 african swi	233	25	100.0	512	1	MATK_LILTS	O9gig3 lilium tain
161	25	100.0	419	1	DNLI_ASEF2	P26813 african swi	234	25	100.0	513	1	AOL1_THBAQ	P08594 thermus aqu
162	25	100.0	420	1	CLPX_BACSU	P50866 bacillus su	235	25	100.0	513	1	YFKQ_BACSU	O34486 bacillus su
163	25	100.0	421	1	CLPX_LACPL	Q89ve2 lactobacill	236	25	100.0	513	1	MATK_MEVDI	O9fmb2 medeola vir
164	25	100.0	422	1	VDR_MOUSE	P48281 mus musculu	237	25	100.0	516	1	NMD3_YEAST	P38861 saccharomyc
165	25	100.0	422	1	VDR_XENLA	O13124 xenopus lae	238	25	100.0	518	1	FX11_MOUSE	Q7cpd1 mus musculu
166	25	100.0	423	1	VDR_RAT	P13053 rattus norv	239	25	100.0	521	1	Y018_METJA	Q0324 methanococc
167	25	100.0	424	1	CLPX_BACHD	Q9k8f4 bacillus ha	240	25	100.0	524	1	TP6B_SULSH	O05207 sulfolobus
168	25	100.0	424	1	VDR_BOVIN	Q85c37 bos taurus	241	25	100.0	530	1	TP6B_SULSO	O97zfo sulfolobus
169	25	100.0	424	1	CLPX_THETN	O8rc24 thermoanaer	242	25	100.0	530	1	PTM1_YEAST	P32857 saccharomyc
170	25	100.0	425	1	SYH1_BACAA	Q8in41 bacillus an	243	25	100.0	531	1	GMT1_APRSP	P21134 aphanocapsa
171	25	100.0	426	1	CLPX_MYCLE	Q9cby6 mycobacteri	244	25	100.0	532	1	GMT1_APRSP	P21134 aphanocapsa
172	25	100.0	426	1	CLPX_MYCTU	O53184 mycobacteri	245	25	100.0	535	1	ALGG_PSESM	P46098 hordeum vul
173	25	100.0	426	1	SVH1_BACCR	Q81b71 bacillus ce	246	25	100.0	536	1	V155_FOWPV	Q887q3 pseudomonas
174	25	100.0	427	1	YRVC_LACLE	P48795 lactobacill	247	25	100.0	542	1	V322_MYCGE	O915a7 fowlpox vir
175	25	100.0	427	1	VDR_HUMAN	P11473 homo sapien	248	25	100.0	558	1	SYP_BACSU	P47564 mycoplasma
176	25	100.0	428	1	CLPX_STRCO	Q9f316 streptomyce	249	25	100.0	564	1	HXB1_HAEIN	O31755 bacillus su
177	25	100.0	428	1	H1SX_CHLTE	Q8key6 chlorobium	250	25	100.0	565	1	HXB1_HAEIN	O44601 haemophilus
178	25	100.0	430	1	OSTB_YEAST	P33767 saccharomyc	251	25	100.0	565	1	HXB1_HAEIN	P43556 haemophilus
179	25	100.0	430	1	SRYD_DROME	P07664 drosophila	252	25	100.0	566	1	HEMA_IAZUK	P26141 influenza a



253 25 100.0 566 1 NPPE BACCE  
 254 25 100.0 569 1 MAMC PIRSP  
 255 25 100.0 569 1 YLS3 YEAST  
 256 25 100.0 571 1 MAMC\_PIRSP  
 257 25 100.0 571 1 SYP BUCAP  
 258 25 100.0 572 1 SYP HABIN  
 259 25 100.0 572 1 SYP HELPJ  
 260 25 100.0 577 1 SYP HELPY  
 261 25 100.0 580 1 RUPA CHIRE  
 262 25 100.0 581 1 PRUR BOVIN  
 263 25 100.0 581 1 PRUR CREEL  
 264 25 100.0 581 1 PRUR SHEEP  
 265 25 100.0 589 1 MAYS YEAST  
 266 25 100.0 589 1 MAZS YEAST  
 267 25 100.0 591 1 TYPA ECOLI  
 268 25 100.0 594 1 FPQT HUMAN  
 269 25 100.0 597 1 GLMS PYRFO  
 270 25 100.0 602 1 LEPA CHLMU  
 271 25 100.0 602 1 LEPA CHLTR  
 272 25 100.0 602 1 YH02 YEAST  
 273 25 100.0 605 1 APW2 YEAST  
 274 25 100.0 606 1 MANA\_PIRSP  
 275 25 100.0 607 1 TYPA ECOL6  
 276 25 100.0 608 1 GHR CHICK  
 277 25 100.0 611 1 GHR COLLI  
 278 25 100.0 611 1 LEPA ENTEA  
 279 25 100.0 616 1 HTPG BORBU  
 280 25 100.0 616 1 TYPA HAEIN  
 281 25 100.0 617 1 SYP SYN3  
 282 25 100.0 620 1 SYT FUSNN  
 283 25 100.0 624 1 COGS METKA  
 284 25 100.0 626 1 UVRC SYNEL  
 285 25 100.0 626 1 UVRC SYNY3  
 286 25 100.0 627 1 UVRC ANASP  
 287 25 100.0 631 1 Y282 MYCTU  
 288 25 100.0 634 1 GHR BOVIN  
 289 25 100.0 634 1 GHR SHEEP  
 290 25 100.0 635 1 MERH DROME  
 291 25 100.0 638 1 COGS METKA  
 292 25 100.0 638 1 GHR HUMAN  
 293 25 100.0 638 1 GHR MACMU  
 294 25 100.0 638 1 GHR PIG  
 295 25 100.0 638 1 GHR RABIT  
 296 25 100.0 638 1 GHR RAT  
 297 25 100.0 639 1 DHAE ECOLI  
 298 25 100.0 640 1 Y282 MYCLE  
 299 25 100.0 642 1 SYT BUCAI  
 300 25 100.0 650 1 GHR\_MOUSE

## ALIGNMENTS

RESULT 1  
 Y01K\_BPT4  
 ID Y01K\_BPT4 STANDARD; PRT; 54 AA.  
 AC P39228; Q9TQV8;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical 5.9 kDa protein in Gp58-sp intergenic region.  
 GN Y01K OR 61.1.  
 OS Bacteriophage T4.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 OC T4-like viruses.  
 OC NCBI\_TaxID=10665;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=93188183; PubMed=8383243;  
 RX Sellick H.E., Stormo G.D., Dyson R.L., Alberts B.M.;  
 RT "Analysis of five presumptive protein-coding sequences clustered  
 between the primosome genes, 41 and 61, of bacteriophages T4, T2, and  
 T6.";

RL J. Virol. 67:2305-2316(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22514363; PubMed=12626685;  
 RA Miller B.S., Kutter E., Mosig G., Arita F., Kunisawa T., Ruger W.;  
 RT "Bacteriophage T4 Genome";  
 RL Microbiol. Mol. Biol. Rev. 67:86-156(2003).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; S57514; AAB25712.1; -.  
 DR EMBL; AF158101; AAD42508.1; -.  
 DR PIR; E45681; E45681.  
 KW Hypothetical protein.  
 FT CONFLICT 26 26 I -> T (IN REF. 1).  
 SQ SEQUENCE 54 AA; 5897 MW; 972CS16A70F86609 CRC64;  
 Query Match 100.0%; Score 25; DB 1; Length 54;  
 Best Local Similarity 50.0%; Pred. No. 50;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEPIXXDX 8  
 DB 5 KEFIKEDM 12  
 RESULT 2  
 VES\_HPV13  
 ID VES\_HPV13 STANDARD; PRT; 91 AA.  
 AC Q02267;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Probable E5 protein.  
 GN E5.  
 OS Human papillomavirus type 13.  
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
 OC Papillomavirus.  
 OX NCBI\_TaxID=10573;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92391075; PubMed=1325697;  
 RA van Ranst M., Fuse A., Fiten P., Beuken E., Pfister H., Burk R.D.,  
 RA Opendakker G.;  
 RT "Human papillomavirus type 13 and pygmy chimpanzee papillomavirus  
 type 1: comparison of the genome organizations.";  
 RL Virology 190:587-596(1992).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X62843; CA444652.1; -.  
 DR EMBL; F42955; W5W113.  
 DR InterPro; IPR008736; Papilloma\_E5A.  
 DR Pfam; PF05776; Papilloma\_E5A; 1.  
 KW Early protein.  
 SQ SEQUENCE 91 AA; 10245 MW; FCBOCE86BB7E5DE8 CRC64;  
 Query Match 100.0%; Score 25; DB 1; Length 91;  
 Best Local Similarity 50.0%; Pred. No. 90;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;



```
DR EMBL: K01263; -: NOT_ANNOTATED_CDS.
DR PIR: A03800; Q0SE8H.
KW Hypothetical protein.
SQ SEQUENCE 118 AA; 13412 MW; 3FE7F7A48658E208 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 118;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 28 QEFITNDV 35
:||||:

RESULT 7
NUSB_STAEP STANDARD; PRT; 129 AA.
AC Q8CP37;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE N utilization substance protein B homolog (NusB protein).
GN NUSB OR SE1205.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qiu Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RA "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
CC -!- FUNCTION: Involved in the transcription termination process (By
CC similarity).
CC -!- SIMILARITY: Belongs to the nusB family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE016748; AAC04804.1; -.
DR HMAP; MF_00073; -.
DR InterPro; IPR006027; NusB.
DR Pfam; PF01029; NusB; 1.
KW Transcription termination; Complete proteome.
SQ SEQUENCE 129 AA; 15150 MW; 161923DD8535A1CA CRC64;

Query Match 100.0%; Score 25; DB 1; Length 129;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 30 IEFIKDDH 37
:||||:

RESULT 8
STM2_MOUSE
ID STM2_MOUSE STANDARD; PRT; 146 AA.
AC P83093;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Stromal interaction molecule 2 (Fragment).
GN STM2.

Query Match 100.0%; Score 25; DB 1; Length 129;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 30 IEFIKDDH 37
:||||:

RESULT 5
VETM_NPVOP STANDARD; PRT; 109 AA.
AC Q10307;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ECORI-T site protein ETM homolog.
GN ETM.
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RA "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome.";
RL Virology 229:381-399(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U75930; AAC59051.1; -.
DR Early protein.
KW Early protein.
SQ SEQUENCE 109 AA; 12053 MW; 597E7F7137389454 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 109;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 96 LEPIKLDV 103
:||||:

RESULT 6
Y13K_HCMVA STANDARD; PRT; 118 AA.
ID Y13K_HCMVA
AC P03236;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 13 kDa protein in transforming region.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=84090384; PubMed=6317885;
RA Nelson J.A., Fleckenstein B., Jahn G., Galloway D.A., McDougall J.K.;
RA "Structure of the transforming region of human cytomegalovirus
RT AD169.";
RL J. Virol. 49:109-115(1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSU=Skeletal muscle;
RX MEDLINE=21356314; PubMed=11463338;
RA Williams R.T., Manji S.S.M., Parker N.J., Hancock M.S.,
RA Van Stekelenburg L., Eid J.-P., Senior P.V., Kazenwadel J.S.,
RA Shandala T., Saint R., Smith P.J., Dziadek M.A.;
RT "Identification and characterization of the STIM (stromal interaction
RT molecule) gene family: coding for a novel class of transmembrane
RT proteins.";
RL Biochem. J. 357:673-685(2001).
CC -!- FUNCTION: Possible adhesion molecule with a role in early
CC hematopoiesis by mediating attachment to stromal cells. Influences
CC the survival and/or proliferation of B cell precursors. Binding to
CC cells requires Mn(II) (By similarity).
CC -!- SUBUNIT: Oligomer with STIM1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell surface
CC (Potential).
CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF328907; AAK82339.1; -.
CC MGD; MGI:2151156; Stim2.
CC GO; GO:0005515; F:protein binding; IPI.
CC InterPro; IPR001660; SAM.
CC SMART; SM00454; SAM; 1.
CC PROSITE; PS50105; SAM DOMAIN; 1.
CC Cell adhesion; Transmembrane; Coiled coil; Glycoprotein.
CC NON_TER 1 1
CC CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC DOMAIN 55 123 SAM.
CC NON_TER 146 146
CC SEQUENCE 146 AA; 17321 MW; 5037405333EBOF63 CRC64;
Query Match 100.0%; Score 25; DB 1; Length 146;
Best Local Similarity 50.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Qy 1 XEFIXDX 8
Db 12 DEFIREDM 19
RT
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum.";
RL Nature 407:757-762(2000).
CC -!- FUNCTION: Has an important function as a repair enzyme for
CC proteins that have been inactivated by oxidation. Catalyzes the
CC reversible oxidation-reduction of methionine sulfoxide in proteins
CC to methionine (By similarity).
CC -!- CATALYTIC ACTIVITY: Protein L-methionine + oxidized thioredoxin =
CC protein L-methionine S-oxide + reduced thioredoxin.
CC -!- SIMILARITY: Belongs to the msrA Met sulfoxide reductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002125; AAF30698.1; -.
CC DR HSSP; P54149; 1FVG.
CC DR HAMAP; MF_01401; -.
CC DR InterPro; IPR002569; PMSR.
CC Pfam; PF01625; PMSR; 1.

```

```

RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -!- SIMILARITY: Contains 1 response regulatory domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000080; AAB91730.1; -.
CC PIR; T28645; T28645.
CC InterPro; IPR001789; Response_reg.
CC SMART; SM00448; REC; 1.
CC PROSITE; PS50110; RESPONSE_REGULATORY; 1.
CC Hypothetical protein; Plasmid.
CC DOMAIN 1 118 RESPONSE REGULATORY.
CC SEQUENCE 151 AA; 17262 MW; F8B42BCF7845E03F CRC64;
Query Match 100.0%; Score 25; DB 1; Length 151;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Qy 1 XEFIXDX 8
Db 47 PEFILDM 54
RT
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum.";
RL Nature 407:757-762(2000).
CC -!- FUNCTION: Has an important function as a repair enzyme for
CC proteins that have been inactivated by oxidation. Catalyzes the
CC reversible oxidation-reduction of methionine sulfoxide in proteins
CC to methionine (By similarity).
CC -!- CATALYTIC ACTIVITY: Protein L-methionine + oxidized thioredoxin =
CC protein L-methionine S-oxide + reduced thioredoxin.
CC -!- SIMILARITY: Belongs to the msrA Met sulfoxide reductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002125; AAF30698.1; -.
CC DR HSSP; P54149; 1FVG.
CC DR HAMAP; MF_01401; -.
CC DR InterPro; IPR002569; PMSR.
CC Pfam; PF01625; PMSR; 1.

```

```

DR ProDom; PD003489; PMSR; 1.
DR TIGRFAMS; TIGR00401; msrA; 1.
KW Oxidoreductase; Complete proteome.
FT ACT SITE 11 BY SIMILARITY.
SQ SEQUENCE 165 AA; 19341 MW; 678DC470BBE22AAD CRC64;

Query Match 100.0%; Score 25; DB 1; Length 165;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 105 DEFIKDF 112

RESULT 11
ALL3_RHIME STANDARD; PRT; 170 AA.
ID ALL3_RHIME STANDARD; PRT; 170 AA.
AC Q92UG4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ureidoglycolate hydrolase 1 (EC 3.5.3.19).
OS ALHA1 OR RB1167 OR SBE20873.
GN Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=2139508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Gouzy J.,
RA Golding B., Puhler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).
CC -!- CATALYTIC ACTIVITY: (-)-ureidoglycolate + H(2)O = glyoxylate + 2
CC NH(3) + CO(2).
CC -!- PATHWAY: Degradation of allantoin (purine catabolism); third step.
CC -!- SIMILARITY: Belongs to the ureidoglycolate hydrolase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL603646; CAC49567.1; -.
CC FIR; G95987; G95987.
CC HAMAP; MF_00616; -.
CC InterPro; IPR007247; Ureidoglyc_hydro.
CC Pfam; PF04115; Ureidoglyc_hydro; 1.
KW Hydrolase; Purine metabolism; Plasmid; Complete proteome.
SQ SEQUENCE 170 AA; 18930 MW; 051CACE71CBF9CDF CRC64;

Query Match 100.0%; Score 25; DB 1; Length 170;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 136 SEPIVVD 143

RESULT 12
PHCB_SYNP2 STANDARD; PRT; 172 AA.
ID PHCB_SYNP2
AC P03944;

23-OCT-1986 (Rel. 02, Created)
23-OCT-1986 (Rel. 02, Last sequence update)
01-JUL-1993 (Rel. 26, Last annotation update)
DE C-phycocyanin beta chain.
GN PCPC.
OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32049;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PR-6;
RX MEDLINE=85088525; PubMed=6096868;
RA de Lorimier R., Bryant D.A., Porter R.D., Liu W.-Y., Jay E.,
RA Stevens S.E. Jr.;
RT "Genes for the alpha and beta subunits of phycocyanin."
RL Proc. Natl. Acad. Sci. U.S.A. 81:7946-7950 (1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85063716; PubMed=6438628;
RA Pilot T.O., Fox J.L.;
RT "Cloning and sequencing of the genes encoding the alpha and beta
RT subunits of C-phycocyanin from the cyanobacterium Agmenellum
RT quadruplicatum."
RL Proc. Natl. Acad. Sci. U.S.A. 81:6983-6987 (1984).
RN [3]
RP METHYLATION.
RX MEDLINE=88087010; PubMed=2447072;
RA Klotz A.V., Glazer A.N.;
RT "Gamma-N-methylasparagine in phycobilliproteins. Occurrence, location,
RT and biosynthesis."
RL J. Biol. Chem. 262:17350-17355 (1987).
RN [4]
RP TERTIARY STRUCTURE AROUND METHYLATED SITE.
RA Duerrring M., Huber R., Bode W.;
RT "The structure of gamma-N-methylasparagine in C-phycocyanin from
RT Mastogladius laminosus and Agmenellum quadruplicatum."
RL FEBS Lett. 236:167-170 (1988).
CC -!- FUNCTION: Light-harvesting photosynthetic bile pigment-protein
CC from the phycobilliprotein complex.
CC -!- SUBUNIT: Heterodimer of an alpha and a beta chain.
CC -!- PTM: Contains two covalently linked bilin chromophores.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; K02660; AAB05343.1; -.
CC EMBL; K02659; AAB05341.1; -.
CC HSSP; P07119; 1CPC.
CC InterPro; IPR001659; Phycobilisome.
CC InterPro; IPR006247; Phycocyanin_b.
CC Pfam; PF00502; Phycobilisome; 1.
CC ProDom; PD000340; Phycobilisome; 1.
DR TIGRFAMS; TIGR01339; phycocy_beta; 1.
KW Phycobilisome; Electron transport; Photosynthesis; Bile pigment;
KW Methylation.
FT MOD_RES 72 72 METHYLATION.
FT BINDING 82 82 PHYCOCYANOBILIN CHROMOPHORE.
FT BINDING 153 153 PHYCOCYANOBILIN CHROMOPHORE.
SQ SEQUENCE 172 AA; 18336 MW; 70DB2718A7767EC CRC64;

Query Match 100.0%; Score 25; DB 1; Length 172;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 16 GEFISDK 23

```

```

RESULT 13
ID AROK_VIBCH STANDARD; PRT; 174 AA.
AC Q9KNV1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Shikimate kinase (EC 2.7.1.71) (SK).
GN AROK OR VC2629.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN SEQUENCE FROM N.A.
RP STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolova M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -!- CATALYTIC ACTIVITY: ATP + shikimate = ADP + shikimate 3-phosphate.
CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC fifth step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the shikimate kinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE004329; AAF95770.1; -.
CC HSP; P10880; 1SKH.
CC TIGR; VC2629; -.
CC HAWAP; MF_00109; -.
CC InterPro; IPR000623; Shik_kinase.
CC Pfam; PF01202; SKI; 1.
CC PRINTS; PR01100; SHIKIMTKINASE.
CC PROSITE; PS01128; SHIKIMATE KINASE; 1.
KW Aromatic amino acid biosynthesis; Transferase; Kinase; ATP-binding;
FT NP BIND 11 18 ATP (BY SIMILARITY).
SQ SEQUENCE 174 AA; 19618 MW; 206FABCBA9981CB CRC64;

Query Match 100.0%; Score 25; DB 1; Length 174;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 XEFIXXDX 8
DB 97 IEFIXKDI 104
:||||:
:||||:

RESULT 15
PSAF ODOSI STANDARD; PRT; 185 AA.
AC P49483;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Photosystem I reaction centre subunit III (PSI-F).
GN PSAF.
OS Odontella sinensis (Marine centric diatom).
OG Chloroplast.
OC Eukaryota; Stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC Biddulphiophycidae; Eupodiscaceae; Eupodiscaceae; Odontella.
OX NCBI_TaxID=2839;
RN [1]
RP SEQUENCE FROM N.A.
RA Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
RT "The chloroplast genome of a chlorophyll a+c-containing alga,
RT Odontella sinensis."
RL Plant Mol. Biol. Rep. 13:336-342(1995).
CC -!- FUNCTION: Probably participates in efficiency of electron transfer
CC from plastocyanin to P700 (or cytochrome c553 in algae and
CC cyanobacteria). This plastocyanin-docking protein contributes to
CC the specific association of plastocyanin to PSI.
CC -!- SUBCELLULAR LOCATION: Associated with luminal side of the
CC thylakoid membrane.
CC -!- SIMILARITY: Belongs to the psaf family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; Z67753; CAA91702.1; -.

DR PIR; S78329; S78329.

DR InterPro; IPR003666; PSI\_PsaF.

DR Pfam; PF02507; PSI\_PsaF; 1.

KW Photosynthesis; Photosystem I; Chloroplast; Thylakoid; Membrane.

SQ SEQUENCE 185 AA; 20747 MW; 16C70038FAD4F048 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 185;

Best Local Similarity 50.0%; Pred. No. 2e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 XEFIXDX 8

Db 166 QEFISNDL 173

RESULT 16

ID Y001 BORBU

AC O51035; O51894; PRT; 190 AA.

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein B80001.

GN B80001.

OS Borrelia burgdorferi (Lyme disease spirochete).

OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.

OX NCBI\_TaxID=139;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35210 / B31;

RX MEDLINE=98065943; PubMed=9403685;

RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,

RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,

RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,

RA Peterson J., Karlvage A.R., Quackenbush J., Salzberg S., Hanson M.,

RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,

RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,

RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,

RA Smith H.O., Venter J.C.;

RT "Genomic sequence of a Lyme disease spirochaete, Borrelia

RT burgdorferi";

RL Nature 390:580-586 (1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Sh-2-82;

RX MEDLINE=98065594; PubMed=9402027;

RA Casjens S., Murphy M., Delange M., Sampson L., van Vugt R.,

RA Huang W.M.;

RT "Telomeres of the linear chromosomes of Lyme disease spirochaetes:

RT nucleotide sequence and possible exchange with linear plasmid

RT telomeres";

RL Mol. Microbiol. 26:581-596 (1997).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AE001115; AAC66406.1; -.

DR EMBL; AF008218; AAB93996.1; -.

DR PIR; A70100; A70100.

DR TIGR; B80001; -.

KW Hypothetical protein; Complete proteome.

FT CONFLICT 12 12 I -> M (IN REF. 2).

FT CONFLICT 34 36 EDL -> GDI (IN REF. 2).  
 FT CONFLICT 40 41 KI -> NL (IN REF. 2).  
 FT CONFLICT 53 53 S -> P (IN REF. 2).  
 FT CONFLICT 81 81 S -> I (IN REF. 2).  
 SQ SEQUENCE 190 AA; 22491 MW; FBFCE799CD827B45 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 190;

Best Local Similarity 50.0%; Pred. No. 2e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 XEFIXDX 8

Db 177 KERILDL 184

RESULT 17

ID GLPP\_BACSU

AC P30300; STANDARD; PRT; 192 AA.

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE GlyceroL uptake operon antiterminator regulatory protein.

GN GLPP OR BSU09270.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93171878; PubMed=8436953;

RA Beijer L., Nilsson R.-P., Holmberg C., Rutberg L.;

RT "The glpP and glpF genes of the glycerol regulon in Bacillus

RT subtilis";

RL J. Gen. Microbiol. 139:349-359 (1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98240224; PubMed=9579061;

RA Noback M.A., Holsappel S., Kiewiet R., Terpstra P., Wambutt R.,

RA Wedler H., Venema G., Bron S.;

RT "The 172 kb pKRA- $\alpha$ addAB region from 83 degrees to 97 degrees of the

RT Bacillus subtilis chromosome contains several dysfunctional genes,

RT the glpB marker, many genes encoding transporter proteins, and the

RT ubiquitous hit gene";

RL Microbiology 144:859-875 (1998).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,

RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,

RA Presecan E., Pujic P., Farnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,

RA Sato T., Scanlan E., Schleicher S., Schroeter R., Scoffone P.,

RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,

```
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256 (1997).
CC -1- FUNCTION: IN THE PRESENCE OF GLYCEROL 3-PHOSPHATE, THIS PROTEIN
CC IS THOUGHT TO ENHANCE TRANSCRIPTION.
CC -1- SIMILARITY: TO E.COLI YGPK.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M99611; AAA22489.1; -
CC EMBL; Y14079; CAA74427.1; -
CC EMBL; Z99108; CAB12755.1; -
CC PIR; B47700; B47700.
CC Subtilisin; BG10185; glpp.
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR006699; G3P_antiterm.
CC Pfam; PF04309; G3P_antiterm; 1.
KW Glycerol metabolism; Transcription regulation; Complete proteome.
SQ SEQUENCE 192 AA; 21609 MW; E3B9A85C5AFA3949 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 192;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 71 AEFICQDI 78

RESULT 18
Y4DX RHISN
ID Y4DX RHISN STANDARD; PRT; 192 AA.
AC P55423;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 20.6 kDa protein Y4DX.
GN Y4DX.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: None obvious.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB000070; AAB92444.1; -
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 192 AA; 20638 MW; A2B0B751AD579EB5 CRC64;
```

```
Query Match 100.0%; Score 25; DB 1; Length 192;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 118 VERIEDDP 125

RESULT 19
YG5C YEAST
ID YG5C YEAST STANDARD; PRT; 196 AA.
AC P53317;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 22.3 kDa protein in MGAI-GCN4 intergenic region.
GN YGR251W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / FY1679;
RX MEDLINE=97279234; PubMed=9133742;
RA Feroli F., Carignani G., Favanello A., Guerreiro P., Azevedo D.,
RA Rodrigues-Pousada C., Melchiorretto P., Panzeri L.,
RA Agostoni Carbone M.L.;
RT "Analysis of a 17.9 kb region from Saccharomyces cerevisiae
RT chromosome VII reveals the presence of eight open reading frames,
RT including BRP1 (TFIIB70) and GCN5 genes.";
RL Yeast 13:373-377(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z73036; CAA97280.1; -
CC PIR; S64583; S64583.
CC GeneOnline; 141563; -
CC SGD; S0003483; YGR251W.
KW Hypothetical protein.
SQ SEQUENCE 196 AA; 22320 MW; 9D106ADCA2BF018E CRC64;

Query Match 100.0%; Score 25; DB 1; Length 196;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 93 GEFIEGDK 100

RESULT 20
YK6K ECOLI
ID YK6K ECOLI STANDARD; PRT; 196 AA.
AC P71301;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yk6k.
GN YK6K OR B0294.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
```



```

RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, TO E.COLI YQEH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; U73857; AAB18023.1; -.
DR EMBL; AE000137; AAC73397.1; -.
DR PIR; F64755; F64755.
DR EcoGene; EG14324; ykgK.
DR InterPro; IPR000792; HTH_LuxR.
DR Pfam; PF00196; GerE; 1.
DR PRINTS; PR00038; HTHLUXR.
DR ProDom; PD000307; HTH_LuxR; 1.
DR SMART; SM00421; HTH_LUXR; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 196 AA; 23274 MW; 5A3676E1E79793B9 CRC64;
SQ
Query Match 100.0%; Score 25; DB 1; Length 196;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXXXD 8
DB 72 SEFINRDN 79
RESULT 21
ID IF3_MYCPN STANDARD; PRT; 201 AA.
AC F78024;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Translation initiation factor IF-3.
GN IF3 OR MPN115 OR MP039.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- FUNCTION: IF-3 binds to the 30S ribosomal subunit and shifts the
CC equilibrium between 70S ribosomes and their 50S and 30S subunits in
CC favor of the free subunits, thus enhancing the availability of 30S
CC subunits on which protein synthesis initiation begins.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

```

```

CC -!- SIMILARITY: Belongs to the IF-3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; AE000004; AAG34732.1; -.
DR PIR; S73365; S73365.
DR HSP; P03000; ITIF.
DR HAMAP; MF 00080; -.
DR InterPro; IPR001288; IF3.
DR Pfam; PF00707; IF3_C; 1.
DR Pfam; PF05198; IF3_N; 1.
DR ProDom; PD002880; IF3; 1.
DR TIGRFAMs; TIGR00168; infc; 1.
DR PROSITE; PS00938; IF3; 1.
KW Initiation factor; Protein biosynthesis; Complete proteome.
SQ SEQUENCE 201 AA; 23145 MW; 3E14B9FA797C80C8 CRC64;
Query Match 100.0%; Score 25; DB 1; Length 201;
Best Local Similarity 50.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXXXD 8
DB 44 NEFILIDE 51
RESULT 22
ID DCD_METUA STANDARD; PRT; 204 AA.
AC Q57872;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bifunctional deaminase/diphosphatase (MjPCD-DUT) (PCD/DUT) [Includes:
DE Deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP deaminase);
DE Deoxyuridine 5'-triphosphate nucleotidohydrolase (EC 3.6.1.23)
DE (dUTPase) (dUTP pyrophosphatase)].
GN DCD OR MJ0430.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2861 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen A.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
RN [2]
RP CHARACTERIZATION, MASS SPECTROMETRY, AND MUTAGENESIS OF ASP-135 AND
RP GLU-145.
RX MEDLINE=22538493; PubMed=12538648;
RA Li H., Xu H., Graham D.E., White R.H.;
RT "The Methanococcus jannaschii dCTP deaminase is a bifunctional
RT deaminase and diphosphatase.";
RL J. Biol. Chem. 278:11100-11106(2003).
CC -!- FUNCTION: Catalyzes two consecutive reactions to form dUMP using
CC dCTP as substrate.

```



01-FEB-1995 (Rel. 31, Last sequence update)  
10-OCT-2003 (Rel. 42, Last annotation update)  
Uracyl phosphoribosyltransferase (EC 2.4.2.9) (UMP pyrophosphorylase)  
(UPRTase).  
GN UPP OR IPC-35D OR BSU36890.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=95095982; PubMed=7798145;  
RA Martinussen J., Glaser P., Andersen P.S., Saxild H.H.;  
RT "Two genes encoding uracyl phosphoribosyltransferase are present in  
RT Bacillus subtilis."  
RL J. Bacteriol. 177:271-274 (1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaubner-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Maue J., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche E., Rose M., Sadale Y.,  
RA Sato T., Scanlan E., Scheich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Taconci E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Takashi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis."  
RL Nature 390:249-256 (1997).  
CC -!- CATALYTIC ACTIVITY: UMP + diphosphate = uracyl + 5-phospho-  
CC alpha-D-ribose 1-diphosphate.  
CC -!- COFACTOR: Magnesium (by similarity).  
CC -!- PATHWAY: Pyrimidine salvage pathway.  
CC -!- SIMILARITY: Belongs to the UPRTase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z38002; CAAB6111.1; -;  
CC DR EMBL; Z39122; CAB15706.1; -;  
CC DR FIR; I40484; I40484.  
CC DR HSSP; Q26998; 1BD3.  
CC DR Subtilist; BG10945; upp.  
CC DR HAMAP; MF\_01218; -; 1.  
CC DR InterPro; IPR000836; PRtransferase.  
CC DR InterPro; IPR005765; Ura\_phospho\_trans.

DR Pfam; PF00156; Pribosyltran; 1.  
DR TIGRFAMs; TR001091; upp; 1.  
KW Transferase; Glycosyltransferase; Magnesium; Complete proteome.  
FT DOMAIN 127 139 PRPP-BINDING (BY SIMILARITY).  
GN DOMAIN 193 201 UMP-BINDING, URACYL PART (BY SIMILARITY).  
SQ SEQUENCE 209 AA; 23037 MW; 457C6810A3651564 CRC64;  
Query Match 100.0%; Score 25; DB 1; Length 209;  
Best Local Similarity 50.0%; Pred. No. 2.3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 1 XEFIXDX 8  
DB 125 REFVWDP 132  
RESULT 26  
KTHY\_SCHPO STANDARD; PRT; 210 AA.  
AC P36590; O74528;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Thymidylate kinase (EC 2.7.4.9) (dTMP kinase).  
GN TMP OR SCC70.07C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93003330; PubMed=1327149;  
RA Abaigar L.T., Yeh Y.I., Jong A.Y.;  
RT "Functional and structural conservation of Schizosaccharomycetes pombe  
RT dTMP kinase gene."  
RL Biochim. Biophys. Acta 1132:222-224 (1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Scours J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeall J.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Wolkaert G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Mostl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomycetes pombe."  
RL Nature 415:871-880 (2002).  
CC -!- FUNCTION: Catalyzes the conversion of dTMP to dTDP.  
CC -!- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate = ADP + thymidine  
CC 5'-diphosphate.  
CC -!- PATHWAY: Biosynthesis of dTMP from dTMP.  
CC -!- SIMILARITY: Belongs to the thymidylate kinase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X65868; CAA46698.1; --  
CC EMBL; AL023794; CAA19357.1; --  
CC PIR; S28955; S28955.  
CC PIR; T41553; T41553.  
CC HSP; P00572; 1TWK.  
CC GeneDB Spombe; SPCC70.07c; --  
CC InterPro; IPR000062; Thymidylate\_kin.  
CC Pfam; PF02223; thymidylate\_kin; 1.  
CC TIGRFAMS; TIGR00041; DTMP\_kinase; 1.  
CC PROSITE; PS01331; THYMIDYLATE\_KINASE; 1.  
CC Transferrase; Kinase; Nucleotide biosynthesis; ATP-binding.  
CC NP\_BIND 14 21 ATP (POTENTIAL).  
CC CONFLICT 33 39 SQHEKAE -> LNMKRLK (IN REF. 1).  
CC CONFLICT 59 59 K -> T (IN REF. 1).  
CC CONFLICT 80 93 TIQIYEIQINGVT -> PSIIYRANQQRN (IN REF.  
CC 1).  
CC CONFLICT 125 125 P -> T (IN REF. 1).  
CC CONFLICT 164 164 F -> L (IN REF. 1).  
CC CONFLICT 186 186 S -> YA (IN REF. 1).  
CC CONFLICT 191 191 H -> D (IN REF. 1).  
CC SEQUENCE 210 AA; 24249 MW; 4266144AEDAB68C0 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 210;  
Best Local Similarity 50.0%; Pred. No. 2.3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDXD 8

Db 176 LEFITLDA 183

#### RESULT 27

ID URE1 HELMU STANDARD; PRT; 213 AA.  
AC P50044;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Urease alpha subunit (EC 3.5.1.5) (Urea amidohydrolase) (Fragment).  
GN UREA.  
OS Helicobacter mustelae.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Helicobacteraceae; Helicobacter.  
OX NCBI\_TaxID=217;

SEQUENCE FROM N.A.

RA Solnick J.V., Josenhans C., Tompkins L.S., Labigne A.;  
RT "Construction and characterization of an isogenic urease-negative  
RL mutant of Helicobacter mustelae."  
RL Infect. Immun. 63:3718-3721(1995).  
CC -!- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).  
CC -!- SUBUNIT: CONSIST OF TWO SUBUNITS (ALPHA AND BETA).  
CC -!- CAUTION: IN HELICOBACTER THE ALPHA SUBUNIT IS WHAT IS KNOWN, IN  
CC OTHER BACTERIA AS THE BETA SUBUNIT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L33462; AAC41483.1; --  
DR HSP; P41022; IUBP.  
DR HAMAP; MF 00739; fused; 1.  
DR InterPro; IPR002019; Urease\_beta.  
DR InterPro; IPR002026; Urease\_gamma.  
DR InterPro; IPR008223; Urease\_gamma\_beta.  
DR Pfam; PF00699; Urease\_beta; 1.  
DR Pfam; PF00547; Urease\_gamma; 1.  
DR PIRSF; PIRSF001225; Urease\_gamma\_beta; 1.  
DR ProDom; PD002326; Urease\_beta; 1.  
DR ProDom; PD002319; Urease\_gamma; 1.  
DR TIGRFAMS; TIGR00192; Urease\_beta; 1.  
DR TIGRFAMS; TIGR00193; Urease\_gamma; 1.  
KW Hydrolase.  
FT NON TER 1  
SQ SEQUENCE 213 AA; 23794 MW; B97A6A3FD2CB2F40 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 213;  
Best Local Similarity 50.0%; Pred. No. 2.3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDXD 8

Db 97 GEFILKDE 104

#### RESULT 28

GTXL SOLTU STANDARD; PRT; 217 AA.  
ID GTXL SOLTU STANDARD; PRT; 217 AA.  
AC P32111;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Probable glutathione S-transferase (EC 2.5.1.18) (Pathogenesis-related  
DE protein 1).  
DE PRF1.  
OS Solanum tuberosum (Potato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC lamids; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4113;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92063197; PubMed=2132026;

RA Taylor J.L., Fritzenmeyer K.H., Hauser I., Kombrink E., Rohwer F.,  
RA Schroeder M., Strittmatter G., Hahlbrock K.;  
RT "Structural analysis and activation by fungal infection of a gene  
RT encoding a pathogenesis-related protein in potato."  
RL Mol. Plant Microbe Interact. 3:72-77(1990).

CC -!- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.  
CC -!- INDUCTION: By fungal infection.

CC -!- SIMILARITY: Belongs to the GST superfamily. HSP26 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; J03679; AAA68430.1; --

DR PIR; T07595; T07595.

DR InterPro; IPR004046; GST\_Cterm.

DR InterPro; IPR004045; GST\_Nterm.

DR Pfam; PF00043; GST\_C; 1.

DR Pfam; PF02798; GST\_N; 1.

KW Plant defense; Transferase; Pathogenesis-related protein.

DR TIGRFAMS; TIGR00193; Urease\_gamma; 1.

SQ SEQUENCE 217 AA; 25056 MW; 7B0DBBE216685B4E CRC64;

Query Match

Best Local Similarity 100.0%; Score 25; DB 1; Length 217;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 XEFIXDX 8
Db 29 YEFIEDL 36

RESULT 29
PYRH ARCFU
ID PYRH ARCFU STANDARD; PRT; 219 AA.
AC O28237;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable uridylylate kinase (EC 2.7.4.-) (UK) (Uridine monophosphate
DE kinase) (UMP kinase).
GN PYRH OR AF2042.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervilave A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -!- FUNCTION: Catalyzes the phosphorylation of UMP to UDP (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + UMP = ADP + UDP.
CC -!- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; first
CC step.
CC -!- SIMILARITY: Belongs to the UMP kinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; A5000962; AAB89213.1; -
DR PIR; A69505; A69505.
DR TIGR; AF2042; -.
DR HAMAP; MF 01220; -; 1.
DR InterPro; IPR001048; Aa_kinase.
DR Pfam; PF00696; aakinese; 1.
KW Transferase; Kinase; Pyrimidine biosynthesis; Complete proteome.
SQ SEQUENCE 219 AA; 2339 MW; 751B894ED5261B91 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 219;
Best Local Similarity 50.0%; Pred. NO. 2.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 124 AEFKADV 131

RESULT 30
VE26_NPVAC
ID VE26_NPVAC STANDARD; PRT; 225 AA.
AC P12827;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Early 25.9 kDa protein.
GN DA26.
OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E2;
RX MEDLINE=88091055; PubMed=2826808;
RA Guarino L.A., Summers M.D.;
RT "Functional mapping of Autographa californica nuclear polyhedrosis
RT virus genes required for late gene expression."
RL J. Virol. 62:463-471(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=L1;
RX MEDLINE=90264832; PubMed=2189022;
RA O'Reilly D.R., Passarelli A.L., Goldman I.F., Miller L.K.;
RT "Characterization of the DA26 gene in a hypervariable region of the
RT Autographa californica nuclear polyhedrosis virus genome."
RL J. Gen. Virol. 71:1029-1037(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=E2;
RX MEDLINE=93079853; PubMed=1333113;
RA Braumagel S.C., Daniel K.D., Reilly L.M., Guarino L.A., Hong T.,
RA Summers M.D.;
RT "Sequence, genomic organization of the EcoRI-A fragment of Autographa
RT californica nuclear polyhedrosis virus, and identification of a
RT viral-encoded protein resembling the outer capsid protein VP8 of
RT rotavirus."
RL Virology 191:1003-1008(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RX MEDLINE=94303173; PubMed=8030224;
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus."
RL Virology 202:586-605(1994).
CC -!- FUNCTION: This protein is required for viral late gene expression.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M22619; AAA69846.1; -
DR EMBL; M18857; AAA66808.1; -
DR EMBL; M96361; AAA66786.1; -
DR EMBL; L22858; AAA66646.1; -
DR PIR; A29891; MNV29.
KW Early protein.
SQ SEQUENCE 225 AA; 25910 MW; D27965D85B2C7980 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 225;
Best Local Similarity 50.0%; Pred. NO. 2.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 195 FEFINPE 202

```

```
RESULT 31
PYRH_METUJA STANDARD; PRT; 226 AA.
AC Q58656;
DT 01-NOV-1997 (Rel. 35, Last Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable uridylylate kinase (EC 2.7.4.-) (UK) (Uridine monophosphate
DE kinase) (UMP kinase).
OS PYRH OR MJ1259.
GN Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OC NCBI_TaxID=2190;
RN [1]
SEQUENCE FROM N.A.
STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
MEDLINE=96337999; PubMed=868087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii."
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Catalyzes the phosphorylation of UMP to UDP (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + UMP = ADP + UDP.
CC -!- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; first
CC step.
CC -!- SIMILARITY: Belongs to the UMP kinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67566; AAB999262.1; ALT_INIT.
CC TIGR; MJ1259; -.
DR HAMAP; MF 01220; -; 1.
DR InterPro; IPR001048; Aa kinase.
DR Pfam; PF00696; aakinasel 1.
DR Transferase; Kinase; Pyrimidine biosynthesis; Complete proteome.
KW SEQUENCE 226 AA; 24099 MW; 7052087EE2AE240D CRC64;
Query Match 100.0%; Score 25; DB 1; Length 226;
Best Local Similarity 50.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db 130 AEFINADL 137
:||||:
:||||:
:||||:
RESULT 32
ORN_ARATH STANDARD; PRT; 227 AA.
AC Q92VE0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable oligoribonuclease (EC 3.1.-.-).
OS AT2G26970 OR T20P8.2.
GN Arabidopsis thaliana (Mouse-ear cress).
OC Arabidopsids thaliana (Streptophyta; Embryophyta; Tracheophyta;
OC Eukaryota; Viridiplantae; Magnoliophyta; eudicotyledons; rosids;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC
euroids II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
STRAIN=cv. Columbia;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Rensing C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Freuss D.,
RA Nierman W.C., White O., Bisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RA thaliana."
RL Nature 402:761-768(1999).
CC -!- FUNCTION: 3'-to-5' exoribonuclease specific for small
CC oligoribonucleotides (By similarity).
CC -!- SIMILARITY: Belongs to the oligoribonuclease family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AC005623; AAC77855.1; -.
CC PIR; B84667; B84667.
DR InterPro; IPR006055; Exonuclease.
DR Pfam; PF00929; Exonuclease; 1.
DR SMART; SM00479; EXOIII; 1.
DR Hypothetical protein; Hydrolase; Exonuclease; Nuclease.
FT ACT SITE 148 148 POTENTIAL.
SQ SEQUENCE 227 AA; 25737 MW; 7126263D1E53D80E CRC64;
Query Match 100.0%; Score 25; DB 1; Length 227;
Best Local Similarity 50.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db 214 LEFISSDI 221
:||||:
:||||:
:||||:
RESULT 33
BLOW_STABP STANDARD; PRT; 228 AA.
AC Q8CTZ3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 6-carboxyhexanoate--CoA ligase (EC 6.2.1.14) (Pimeloyl-CoA synthase).
OS BLOW OR SE0182.
GN Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OC NCBI_TaxID=1282;
RN [1]
SEQUENCE FROM N.A.
STRAIN=ATCC 12228;
RC PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RA "Genome-based analysis of virulence genes in a non-biofilm-forming
RA Staphylococcus epidermidis strain (ATCC 12228)."
RL Mol. Microbiol. 49:1577-1593(2003).
CC -!- FUNCTION: Transforms pimelate into pimeloyl-CoA (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + 6-carboxyhexanoate + CoA = AMP +
CC diphosphate + 6-carboxyhexanoyl-CoA.
CC -!- COFACTOR: Magnesium (By similarity).
CC
```

QY	1	XEFIXDX 8	:   ::	
DB	15	LEFIVDDM 22	:   ::	
RESULT 35				
UBIE_THEME				
ID	UBIE_THEME	STANDARD		
AC	G9WZL2			
DT	10-OCT-2003	(Rel. 42)		
DT	10-OCT-2003	(Rel. 42)		
DT	15-MAR-2004	(Rel. 43)		
DE	Menaquinone biosynthesis			
GN	UBIE_OT TM0753			
OS	Thermotoga maritima			
OC	Bacteria			
OX	NCBI_TaxID=2336			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MSB8 / DSM 313			
RX	MEDLINE=99287316; PubMed			
RA	Nelson K.E., Clayton			
RA	Haft D.H., Hickey E.K.,			
RA	McDonald L., Utterback			
RA	Stewart A.M., Cotton			
RA	Heidelberg J., Sutton			
RA	Salzberg S.L., Smith			
RT	"Evidence for lateral			
RT	genome sequence of <i>Th</i>			
RL	Nature 399:323-329 (19			
CC	!- FUNCTION: Methyl			
CC	dimethylmenaquin			
CC	!- CATALYTIC ACTIVITY			
CC	s-adenosyl-L-homoc			
CC	!- PATHWAY: Menaquin			
CC	!- SIMILARITY: Belon			
CC	THIS SWISS-PROT entry			
CC	between the Swiss In			
CC	the European Bioinfo			
CC	use by non-profit			
CC	modified and this str			
CC	entities requires a l			
CC	or send an email to l			
CC	EMBL; AF001745; AAD35			
DR	PIR; G23373; G72337.			
DR	TIGR; TM0753; -			
DR	HAWAP; MF 01813; -			
DR	InterPro; IPR001601;			
DR	InterPro; IPR000051;			
DR	Pfam; PF01209; Ublm			
DR	PROSITE; PS01183; UBI			
DR	PROSITE; PS01184; UBI			
KW	Menaquinone biosynth			
KW	Complete proteome.			
SQ	SEQUENCE 229 AA; 2			
Query Match				
Best Local Similarity				
Matches 4; Conservat.				
QY	1	XEFIXDX 8	:   ::	
DB	94	GEFIVGDA 101	:   ::	
RESULT 36				
YF07_MYCTU				
ID	YF07_MYCTU			
AC	P71786			
AD	STANDARD			



```
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein RV1507c/MT1555.
GN RV1507C OR MT1555 OR MTCY277.29C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37Rv;
RX MEDLINE=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin T., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback L., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z79701; CAB02021.1; -.
DR EMBL; AE007023; AAK45820.1; -.
DR PIR; F70713; F70713.
DR TIGR; MT1555; -.
DR TubercuList; RV1507c; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 231 AA; 27090 MW; BD18A247DB1B0BF8 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 231;
Best Local Similarity 50.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 25 DEFIYDD 32

RESULT 37
ID RPIA_CHLCV STANDARD; PRT; 233 AA.
AC Q822P7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribose 5-phosphate isomerase A (EC 5.3.1.6) (Phosphoriboisomerase A)
DE (PRI).
GN RPIA OR CCA00632.
OS Chlamydomophila caviae.

OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OC NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GPIC;
RX MEDLINE=22569155; PubMed=12682364;
RX Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,
RA Umayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,
RA Fraser C.M.;
RA "Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC):
RT examining the role of niche-specific genes in the evolution of the
RT Chlamydiaceae.";
RL Nucleic Acids Res. 31:2134-2147 (2003).
CC -!- CATALYTIC ACTIVITY: D-ribose 5-phosphate = D-ribulose 5-phosphate.
CC -!- PATHWAY: Nonoxidative branch of the pentose phosphate pathway.
CC -!- SIMILARITY: Belongs to the ribose 5-phosphate isomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016996; AAP05374.1; -.
DR TIGR; CCA00632; -.
DR HAMAP; MF 00170; -.
DR InterPro; IPR004788; RpiA.
DR ProDom; PD005813; RpiA; 1.
DR Isomerase; Complete proteome.
KW Isomerase; Complete proteome.
SQ SEQUENCE 233 AA; 25544 MW; COEDB3474B20A0C1 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 233;
Best Local Similarity 50.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 77 EEFINTDL 84

RESULT 38
ID VHEL_WCMVM STANDARD; PRT; 236 AA.
AC P09499;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Probable helicase (ORF 2).
OS White clover mosaic virus (strain M) (WCMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flexiviridae;
OC Potexvirus.
OC NCBI_TaxID=12189;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86124202; PubMed=3340527;
RA Forster R.L.S., Bevan M.W., Harbison S.-A., Gardner R.C.;
RT "The complete nucleotide sequence of the potexvirus white clover
RT mosaic virus.";
RL Nucleic Acids Res. 16:291-303 (1988).
CC -!- FUNCTION: THIS PROTEIN MAY BE AN ATP-DEPENDENT HELICASE AND IS
CC REQUIRED FOR REPLICATION OF VIRAL DNA.
CC -!- SIMILARITY: TO ORF2 FROM OTHER POTEVIRUSES AND TO 25 kDa PROTEIN
CC FROM CARLAVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; X06728; CAA25905.1; -  
DR PIR; S01086; S01086; Viral helicase.  
DR InterPro; IPR000606; Viral helicase.  
DR Pfam; PF01443; Viral helicase; 1.  
KW DNA replication; ATP-binding; Helicase.  
FT NP BIND 28 35 ATP (POTENTIAL).  
SQ SEQUENCE 236 AA; 26356 MW; 49991932D8B6D4BB CRC64;

Query Match 100.0%; Score 25; DB 1; Length 236;  
Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db 95 FEFIFTDP 102  
:|||||:

RESULT 39  
VHEL WCMVO STANDARD; PRT; 236 AA.  
AC FL5403;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-OCT-1993 (Rel. 27, Last annotation update)  
DE Probable helicase (ORF 2).  
OS White clover mosaic virus (strain O) (WCMV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flexiviridae;  
OC Potexvirus.  
OX NCBI\_TaxID=121190;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=90281578; PubMed=2353451;  
RA Beck D.L., Forster R.L.S., Bevan M.W., Boxen K.A., Lowe S.C.,  
RA Gardner R.C.;  
RT potexvirus white clover mosaic virus.;  
RL Virology 177:152-158(1990).  
CC -!- FUNCTION: THIS PROTEIN MAY BE AN ATP-DEPENDENT HELICASE AND IS  
CC REQUIRED FOR REPLICATION OF VIRAL DNA.  
CC -!- SIMILARITY: TO ORF2 FROM OTHER POTEXVIRUSES AND TO 25 kDa PROTEIN  
CC FROM CARLAVIRUSES.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X16636; CAA34629.1; -  
DR InterPro; IPR000606; Viral helicase.  
DR Pfam; PF01443; Viral helicase; 1.  
KW DNA replication; ATP-binding; Helicase.  
FT NP BIND 28 35 ATP (POTENTIAL).  
SQ SEQUENCE 236 AA; 26451 MW; D470D75B46E9D418 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 236;  
Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db 95 FEFIFTDP 102  
:|||||:

RESULT 40  
YM08 SULSO STANDARD; PRT; 236 AA.  
ID YM08 SULSO  
AC Q97WK8;

DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Hypothetical UPF0286 protein SSO2208.  
GN SSO2208.  
OS Sulfolobus solfataricus.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus  
OX NCBI\_TaxID=2287;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
RX MEDLINE=21132296; PubMed=11427726;  
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
RA Awayez M.J., Chan-Wether C.C.-Y., Clausen I.G., Curtis B.A.,  
RA De Moors A., Erasmo G., Fletcher C., Gordon P.M.K.,  
RA Heikamp-de Jong I., Jeffries A.C., Koza C.J., Medina N., Peng X.,  
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;  
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
CC -!- SIMILARITY: Belongs to the UPF0286 family.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AE006826; AAK42378.1; -  
DR PIR; C90390; C90390.  
DR HAMAP; MF 00722; -; 1.  
DR InterPro; IPR002793; DUF91.  
DR Pfam; PF01939; DUF91; 1.  
DR ProDom; PD013521; DUF91; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 236 AA; 27264 MW; 6930125F102BBAC8 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 236;  
Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db 224 LEFIRYDI 231  
:|||||:

RESULT 41  
YNN9 YEAST STANDARD; PRT; 240 AA.  
ID YNN9 YEAST  
AC P53915;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Hypothetical 27.7 kDa protein in CPT1-SPC98 intergenic region.  
GN YNL129W OR N1219 OR N1870.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=S288c;  
RX MEDLINE=96109932; PubMed=8619318;  
RA Mallet L., Bussereau F., Jacquet M.;  
RT "A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,  
RT MEP2, CAP/SPV2, NAM9, FKBI/FPRI/RBP1, MOM22 and CPT1, predicts an  
RT adenosine deaminase gene and 14 new open reading frames.";  
RL Yeast 11:1195-1209(1995).  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR ENBL; Z46843; CAA86896.1; -.  
 DR ENBL; Z71405; CAA96011.1; -.  
 DR PIR; S55154; S55154.  
 DR Germline; 143135; -.  
 DR SGD; S0005073; YNL129W.  
 DR InterPro; IPR000764; Uridine kin.  
 DR PRINTS; PR00988; URIDINKINASE.  
 KW Hypothetical protein; ATP-binding.  
 FT NP BIND 13 20 ATP (POTENTIAL).  
 SQ SEQUENCE 240 AA; 27689 MW; F04E99885774CF53 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 240;  
 Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8

:|:|:|:

Db 213 KERINDDD 220

RESULT 42

COX2 NEUCR COX2 NEUCR STANDARD; PRT; 250 AA.  
 ID P00411;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1).

GN COX2 OR OXII.

OS Neurospora crassa.

OC Mitochondrion.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

OX NCBI\_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=84032555; PubMed=6313689;

RA Macino G., Morelli G.;

RT "Cytochrome oxidase subunit 2 gene in Neurospora crassa

mitochondria.";

RL J. Biol. Chem. 258:13230-13235(1983).

RN [2]

RP SEQUENCE OF 1-54 FROM N.A.

RX MEDLINE=83054012; PubMed=6291999;

RA van den Boogaart P., van Dijk S., Agsteribbe E.;

RT "The mitochondrially made subunit 2 of Neurospora crassa cytochrome

a3 is synthesized as a precursor protein.";

RL FEBS Lett. 147:97-100(1982).

RN [3]

RP SEQUENCE OF 1-65 FROM N.A.

RC STRAIN=RL3-8A;

RX MEDLINE=89137935; PubMed=2976009;

RA Almasan A., Mishra N.C.;

RT "Molecular characterization of the mitochondrial DNA of a new stopper

mutant ER-3 of Neurospora crassa.";

RL Genetics 120:935-945(1988).

CC -1- FUNCTION: Cytochrome c oxidase is the component of the respiratory

chain that catalyzes the reduction of oxygen to water. Subunits 1-

3 form the functional core of the enzyme complex. Subunit 2

transfers the electrons from cytochrome c via its binuclear copper

A center to the bimetallic center of the catalytic subunit 1.

CC -1- CATALYTIC ACTIVITY: 4 ferrocycytochrome c + O(2) = 4 ferricytochrome

c + 2 H(2)O.

CC -1- COFACTOR: Copper A.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial

CC inner membrane.  
 CC -1- SIMILARITY: Belongs to the cytochrome c oxidase subunit 2 family.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR ENBL; K00825; AAA31959.2; -.  
 DR ENBL; J01429; ; NOT ANNOTATED\_CDS.  
 DR ENBL; X14681; CAA32813.1; -.  
 DR PIR; A00479; OBN2.  
 DR InterPro; IPR001505; Copper\_CuA.  
 DR InterPro; IPR008972; Cupredoxin.  
 DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
 DR Pfam; PF00116; COX2; 1.  
 DR Pfam; PF02790; COX2\_TM; 1.  
 DR PRINTS; PR01166; CYCOXIDASEII.  
 DR PROSITE; PS000131; Copper\_CuA; 1.  
 DR PROSITE; PS00078; COX2; 1.  
 DR Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;  
 KW Electron transport; Respiratory chain.  
 FT DOMAIN 1 39 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
 FT TRANSMEM 40 59 POTENTIAL.  
 FT DOMAIN 60 76 MITOCHONDRIAL MATRIX (POTENTIAL).  
 FT TRANSMEM 77 97 POTENTIAL.  
 FT DOMAIN 98 250 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
 FT METAL 185 185 COPPER A (PROBABLE).  
 FT METAL 220 220 COPPER A (PROBABLE).  
 FT METAL 224 224 COPPER A (PROBABLE).  
 FT METAL 228 228 COPPER A (PROBABLE).  
 SQ SEQUENCE 250 AA; 28724 MW; 75854A38EF16D506 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 250;  
 Best Local Similarity 50.0%; Pred. No. 2.7e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8

:|:|:|:

Db 137 DEFIEFDS 144

RESULT 43

GPP2 YEAST

ID GPP2 YEAST STANDARD; PRT; 250 AA.

AC P40106;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE (DL)-glycerol-3-phosphatase 2 (EC 3.1.3.-).

GN GPP2 OR HOR2 OR YER062C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RS16;

RA Hirayama T., Maeda T., Saito H., Shinozaki K.;

RT "Cloning and characterization of seven cDNAs for hyperosmolarity-

responsive (HOR) genes of Saccharomyces cerevisiae.";

RL Mol. Gen. Genet. 249:127-138(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RX MEDLINE=97313264; PubMed=9169868;

RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,

RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,

RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,

RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,  
RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,  
RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,  
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.,  
RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome V.";  
RL Nature 387:78-81(1997).  
CC [3]  
RN  
RP PARTIAL SEQUENCE, FUNCTION, AND CHARACTERIZATION.  
RX MEDLINE=96278827; PubMed=8662716;  
RA Norbeck J., Paelman A.-K., Akhtar N., Blomberg A., Adler L.;  
RT "Purification and characterization of two isoenzymes of  
RT Di-glycerol-3-phosphatase from *Saccharomyces cerevisiae*.  
RT Identification of the corresponding GPP1 and GPP2 genes and evidence  
RT for osmotic regulation of Gpp2p expression by the osmosensing  
RT mitogen-activated protein kinase signal transduction pathway";  
RL J. Biol. Chem. 271:13875-13881(1996)  
CC -I- CATALYTIC ACTIVITY: Glycerol 3-phosphate + H(2)O = glycerol +  
CC phosphate.  
CC -I- SUBUNIT: Monomer.  
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -I- INDUCTION: By osmotic stress.  
CC -I- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS DOG1, DOG2, GPP1 AND  
CC GPP2.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; D50469; BAA09058.1; -;  
CC EMBL; U18813; AAB64598.1; -;  
CC PIR; S50565; S50565.  
CC GeneOnline; 139142; -;  
CC SGD; S0000864; HOR2.  
CC GO; GO:0000121; F:glycerol-1-phosphatase activity; IDA.  
CC GO; GO:0005975; P:carbohydrate metabolism; IDA.  
CC GO; GO:0006114; P:glycerol biosynthesis; IMP.  
CC GO; GO:0006970; P:response to osmotic stress; IDA.  
CC InterPro; IPR006402; HAD-SF-IA-V3.  
CC InterPro; IPR005834; Hydrolase.  
CC Pfam; PF00702; Hydrolase; 1.  
CC TIGRFAMs; TIGR01509; HAD-SF-IA-V3; 1.  
CC Hydrolyase; Multigene family.  
CC KW  
CC SEQUENCE 250 AA; 27814 MW; D69F343B10417313 CRC64;  
CC  
CC Query Match 100.0%; Score 25; DB 1; Length 250;  
CC Best Local Similarity 50.0%; Pred. No. 2.7e+02;  
CC Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
CC  
CC QY 1 XEFIXDX 8  
CC Db 233 VEFIFDDY 240  
CC  
CC RESULT 44  
CC RS3\_ALDYE STANDARD; PRT; 250 AA.  
CC AC Q8V563;  
CC DT 28-FEB-2003 (Rel. 41, Created)  
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
CC DE 28-FEB-2003 (Rel. 41, Last annotation update)  
CC DE 30S ribosomal protein S3.  
CC GN RPSC OR RPS3.  
CC OS Alder yellows phytoplasma.  
CC OC Bacteria; Firmicutes; Mollicutes; Achleplasmatales;  
CC OC Achleplasmataceae; Phytoplasma.  
CC OX NCBI\_TaxID=72989;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RX MEDLINE=22140588; PubMed=12144771;  
CC

RA Martini M., Botti S., Marcone C., Marzachi C., Casati P., Bianco P.A.,  
RA Benedetti R., Bertaccini A.;  
RT "Genetic variability among flavescence doree phytoplasmas from  
RT different origins in Italy and France.";  
RL Mol. Cell. Probes 16:197-208(2002).  
CC -I- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
CC in the 70S ribosome, positioning it for translation (By  
CC similarity).  
CC -I- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
CC with proteins S10 and S14 (By similarity).  
CC -I- SIMILARITY: Belongs to the S3P family of ribosomal proteins.  
CC -I- SIMILARITY: Contains 1 KH type-2 domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; AF396936; AAL57324.1; -;  
CC HAMAP; MF 01309; -; 1.  
CC InterPro; IPR004087; KH dom.  
CC InterPro; IPR009019; KH\_prok.  
CC InterPro; IPR004044; KH\_TYPE\_2.  
CC InterPro; IPR001351; Ribosomal\_S3\_C.  
CC InterPro; IPR008282; Ribosomal\_S3\_N.  
CC InterPro; IPR005704; S3\_bact.  
CC Pfam; PF00189; Ribosomal\_S3\_C; 1.  
CC Pfam; PF00417; Ribosomal\_S3\_N; 1.  
CC SMART; SM00322; KH; 1.  
CC TIGRFAMs; TIGR01009; rpsC\_bact; 1.  
CC PROSITE; PS00823; KH\_TYPE\_2; 1.  
CC PROSITE; PS00548; RIBOSOMAL\_S3; 1.  
CC KW Ribosomal protein; RNA-binding; rRNA-binding.  
CC FT DOMAIN 39 111 KH TYPE-2.  
CC SQ SEQUENCE 250 AA; 28367 MW; 719D87C695B4F6B4 CRC64;  
CC  
CC Query Match 100.0%; Score 25; DB 1; Length 250;  
CC Best Local Similarity 50.0%; Pred. No. 2.7e+02;  
CC Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
CC  
CC QY 1 XEFIXDX 8  
CC Db 65 DEFIEIDL 72  
CC  
CC RESULT 45  
CC RS3\_ELMPH STANDARD; PRT; 250 AA.  
CC AC Q8VLE1;  
CC DT 28-FEB-2003 (Rel. 41, Created)  
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
CC DE 28-FEB-2003 (Rel. 41, Last annotation update)  
CC DE 30S ribosomal protein S3.  
CC GN RPSC OR RPS3.  
CC OS Elm witches'-broom phytoplasma, and  
CC OS Elm yellows phytoplasma.  
CC OC Bacteria; Firmicutes; Mollicutes; Achleplasmatales;  
CC OC Achleplasmataceae; Phytoplasma.  
CC OX NCBI\_TaxID=182216, 35774;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RX SPECIES=E.w.phytoplasma, and E.y.phytoplasma; STRAIN=UW, and EY1;  
CC RX MEDLINE=22140588; PubMed=12144771;  
CC RA Martini M., Botti S., Marcone C., Marzachi C., Casati P., Bianco P.A.,  
RA Benedetti R., Bertaccini A.;  
RT "Genetic variability among flavescence doree phytoplasmas from  
RT different origins in Italy and France.";  
RL Mol. Cell. Probes 16:197-208(2002).  
CC -I- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
CC in the 70S ribosome, positioning it for translation (By

CC similarity).

CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex

CC with proteins S10 and S14 (By similarity).

CC -1- SIMILARITY: Belongs to the S3P family of ribosomal proteins.

CC -1- SIMILARITY: Contains 1 KH type-2 domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AF396949; AAL57337.1; -

CC EMBL; AF396938; AAL57326.1; -

CC HAMAP; MF 01309; -; 1.

CC InterPro; IPR004087; KH dom.

CC InterPro; IPR009019; KH prok.

CC InterPro; IPR004044; KH TYPE 2.

CC InterPro; IPR001351; Ribosomal\_S3\_C.

CC InterPro; IPR008282; Ribosomal\_S3\_N.

CC InterPro; IPR005704; S3 bact.

CC Pfam; PF00189; Ribosomal\_S3\_C; 1.

CC Pfam; PF00417; Ribosomal\_S3\_N; 1.

CC SMART; SM00322; KH; 1.

CC TIGRFAWS; TIGR01009; rpsC\_bact; 1.

CC PROSITE; PS08223; KH TYPE 2; 1.

CC PROSITE; PS00548; RIBOSOMAL\_S3; 1.

CC Ribosomal protein; RNA-binding; rRNA-binding.

CC DOMAIN 39 111 KH TYPE-2.

CC SEQUENCE 250 AA; 28455 MW; 7356A7E7819EA1A CRC64;

CC -----

CC Query Match 100.0%; Score 25; DB 1; Length 250;

CC Best Local Similarity 50.0%; Pred. No. 2.7e+02;

CC Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

CC -----

CC QY 1 XEFIXDX 8

CC :|||:|:

CC 65 DEFIELD 72

CC Db

CC -----

CC RESULT 46

CC RS3\_FLADO STANDARD; PRT; 250 AA.

CC AC Q8VL42; Q8VL25; Q8VL43; Q8VS62;

CC DT 28-FEB-2003 (Rel. 41, Created)

CC DT 28-FEB-2003 (Rel. 41, Last sequence update)

CC DT 28-FEB-2003 (Rel. 41, Last annotation update)

CC DE 30S ribosomal protein S3.

CC GN RPS3 OR RPS3.

CC OS Flavescence doree phytoplasma.

CC OC Bacteria; Firmicutes; Mollicutes; Acholeplasmatales;

CC OC Acholeplasmataceae; Phycoplasmata.

CC OX NCBI\_TaxID=131152;

CC RN [1]

CC SEQUENCE FROM N.A.

CC RC STRAIN=AL202, AT4, FD70, PD88, PC2, PNG, PV259, TV46, TV51, and VR2;

CC RX MEDLINE=22140588; PubMed=12144771;

CC RA Martini M., Botti S., Marcone C., Marzachi C., Casati P., Bianco P.A.,

CC RA Benedetti R., Bertaccini A.;

CC RT "Genetic variability among flavescence doree phytoplasmas from

CC RT different origins in Italy and France."

CC RL Mol. Cell. Probes 16:197-208(2002).

CC CC -1- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA

CC in the 70S ribosome, positioning it for translation (By

CC similarity).

CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex

CC with proteins S10 and S14 (By similarity).

CC -1- SIMILARITY: Belongs to the S3P family of ribosomal proteins.

CC -1- SIMILARITY: Contains 1 KH type-2 domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AF396937; AAL57325.1; -

CC EMBL; AF396939; AAL57327.1; -

CC EMBL; AF396940; AAL57328.1; -

CC EMBL; AF396942; AAL57330.1; -

CC EMBL; AF396943; AAL57331.1; -

CC EMBL; AF396945; AAL57333.1; -

CC EMBL; AF396946; AAL57334.1; -

CC EMBL; AF396947; AAL57335.1; -

CC EMBL; AF396948; AAL57336.1; -

CC EMBL; AF396950; AAL57338.1; -

CC EMBL; AF396951; AAL57339.1; -

CC HAMAP; MF 01309; -; 1.

CC InterPro; IPR004087; KH dom.

CC InterPro; IPR009019; KH prok.

CC InterPro; IPR004044; KH TYPE 2.

CC InterPro; IPR001351; Ribosomal\_S3\_C.

CC InterPro; IPR008282; Ribosomal\_S3\_N.

CC InterPro; IPR005704; S3 bact.

CC Pfam; PF00189; Ribosomal\_S3\_C; 1.

CC Pfam; PF00417; Ribosomal\_S3\_N; 1.

CC SMART; SM00322; KH; 1.

CC TIGRFAWS; TIGR01009; rpsC\_bact; 1.

CC PROSITE; PS08223; KH TYPE 2; 1.

CC PROSITE; PS00548; RIBOSOMAL\_S3; 1.

CC Ribosomal protein; RNA-binding; rRNA-binding.

CC DOMAIN 39 111 KH TYPE-2.

CC FT VARIANT 46 46 H -> Y (IN STRAINS AL202, AT4 AND PV259).

CC FT VARIANT 103 103 V -> I (IN STRAINS AL202, AT4, FD70 AND

CC FT VARIANT 118 118 V -> I (IN STRAINS TV46 AND TV51).

CC SQ SEQUENCE 250 AA; 28408 MW; E541C224F164FF5 CRC64;

CC -----

CC Query Match 100.0%; Score 25; DB 1; Length 250;

CC Best Local Similarity 50.0%; Pred. No. 2.7e+02;

CC Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

CC -----

CC QY 1 XEFIXDX 8

CC :|||:|:

CC 65 DEFIELD 72

CC Db

CC -----

CC RESULT 47

CC RS3\_RUBST STANDARD; PRT; 250 AA.

CC AC Q8VS60;

CC DT 28-FEB-2003 (Rel. 41, Created)

CC DT 28-FEB-2003 (Rel. 41, Last sequence update)

CC DT 28-FEB-2003 (Rel. 41, Last annotation update)

CC DE 30S ribosomal protein S3.

CC GN RPS3 OR RPS3.

CC OS Rubus stunt phytoplasma.

CC OC Bacteria; Firmicutes; Mollicutes; Acholeplasmatales;

CC OC Acholeplasmataceae; Phycoplasmata.

CC OX NCBI\_TaxID=72996;

CC RN [1]

CC SEQUENCE FROM N.A.

CC RC STRAIN=RUS;

CC RX MEDLINE=22140588; PubMed=12144771;

CC RA Martini M., Botti S., Marcone C., Marzachi C., Casati P., Bianco P.A.,

CC RA Benedetti R., Bertaccini A.;

CC RT "Genetic variability among flavescence doree phytoplasmas from

CC RT different origins in Italy and France."

CC RL Mol. Cell. Probes 16:197-208(2002).

CC CC -1- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA

CC in the 70S ribosome, positioning it for translation (By

CC similarity).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR EMBL: X15441; CAA33481.1; -
DR InterPro: IPR001505; Copper_CuA.
DR InterPro: IPR008972; Cupredoxin.
DR InterPro: IPR002429; Cyt_c_ox_2.
DR Pfam: PF001116; COX2; 1.
DR Pfam: PF02790; COX2; 1.
DR PRINTS: PR01166; CYCOXIDASEII.
DR PRODOM: PD000131; Copper_CuA; 1.
DR PROSITE: PS00078; COX2; 1.
KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
FT DOMAIN 1 39 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT TRANSMEM 40 59 POTENTIAL.
FT DOMAIN 60 81 MITOCHONDRIAL MATRIX (POTENTIAL).
FT TRANSMEM 82 105 POTENTIAL.
FT DOMAIN 106 252 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT METAL 184 184 COPPER A (PROBABLE).
FT METAL 219 219 COPPER A (PROBABLE).
FT METAL 223 223 COPPER A (PROBABLE).
FT METAL 227 227 COPPER A (PROBABLE).
FT METAL 227 227 COPPER A (PROBABLE).
SQ SEQUENCE 252 AA; 28930 MW; 57F98767F8BDC24C CRC64;

Query Match 100.0%; Score 25; DB 1; Length 252;
Best Local Similarity 50.0%; Pred. No. 2.8e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEPIXXDX 8
Db 137 NEFIEFDS 144

RESULT 50
YQXB BACSU STANDARD; PRT; 256 AA.
ID YQXB BACSU STANDARD; PRT; 256 AA.
AC P28671; 054515;
DT 01-DEC-1992 (Rel. 24, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein yoxB (ORF19+).
GN YQXB OR BSU18520.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst P., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boursier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferraci E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaber-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara K., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Rapoport G., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan B., Schleich S., Schroeter R., Scofield F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
"The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256 (1997).
[2]
RN Sequence of 1-119 from N.A.
RP STRAIN=168;
RX MEDLINE=91192601; PubMed=1849493;
RA Ahn K.S., Wake R.G.;
RT "Variations and coding features of the sequence spanning the
RT replication terminus of Bacillus subtilis 168 and W23 chromosomes.";
RL Gene 98:107-112 (1991).
[3]
RN Sequence of 72-256 from N.A.
RP Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the terC and odnAB loci cloned in a yeast artificial chromosome.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: Z99114; CAB13745.1; -
DR EMBL: M24523; -; NOT_ANNOTATED_CDS.
DR EMBL: AF027868; AAB84418.1; -
DR PIR: B69930; B69930.
DR Subtilist; BG11046; yoxB.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 256 AA; 28452 MW; B9B64B08ECE0AB3F CRC64;

Query Match 100.0%; Score 25; DB 1; Length 256;
Best Local Similarity 50.0%; Pred. No. 2.8e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEPIXXDX 8
Db 48 NEFINRDA 55

RESULT 51
DPHB SULSO STANDARD; PRT; 257 AA.
ID DPHB SULSO STANDARD; PRT; 257 AA.
AC Q97TX8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable diphthine synthase (EC 2.1.1.98) (Diphthamide biosynthesis
DE methyltransferase).
GN DPHB OR SSO0953.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasus G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.I., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840 (2001).
CC -!- FUNCTION: Required for the methylation step in diphthamide

```

CC biosynthesis (By similarity).  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + 2-(3-carboxy-3-aminopropyl)-L-histidine = S-adenosyl-L-homocysteine + 2-[3-carboxy-3-(methylanilino)propyl]-L-histidine.  
 CC -1- PATHWAY: Diphthamide biosynthesis; second step.  
 CC -1- SIMILARITY: Belongs to the diphthine synthase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE006715; AAK41227.1; -.  
 CC DR D90246; D90246.  
 CC DR HAMAP; MF 01084; -. 1.  
 CC DR InterPro; IPR000876; Cor/por\_Mettransf.  
 CC DR InterPro; IPR004551; Dphthn\_synthase.  
 CC DR Pfam; PF00590; TP\_methylase; 1.  
 CC DR TIGRFAMs; TIGR00522; dph5; 1.  
 CC KW Transferase; Methyltransferase; Complete proteome.  
 CC SQ SEQUENCE 257 AA; 28934 MW; 218BE19B593BF01B CRC64;  
 CC  
 CC Query Match 100.0%; Score 25; DB 1; Length 257;  
 CC Best Local Similarity 50.0%; Pred. No. 2.8e+02;  
 CC Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 XEFIXDX 8  
 CC Db 55 REFIDR 62  
 CC  
 CC RESULT 52  
 CC ARG1\_SALTY STANDARD; PRT; 260 AA.  
 CC AC P02911;  
 CC DT 21-JUL-1986 (Rel. 01, Created)  
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 CC DE Lysine-arginine-ornithine-binding periplasmic protein precursor  
 CC DE (Lysine-binding protein).  
 CC GN ARG1 OR STM2355.  
 CC OS Salmonella typhimurium.  
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC OC Enterobacteriaceae; Salmonella.  
 CC OX NCBI\_TaxID=602;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=82082392; PubMed=6273842;  
 CC RA Higgins C.F., Ames G.F.-L.;  
 CC RT "Two periplasmic transport proteins which interact with a common membrane receptor show extensive homology: complete nucleotide sequences";  
 CC RT Proc. Natl. Acad. Sci. U.S.A. 78:6038-6042(1981).  
 CC RN [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 CC RX MEDLINE=21534948; PubMed=11677609;  
 CC RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott C., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;  
 CC RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";  
 CC RT Nature 413:852-856(2001).  
 CC RN [3]  
 CC RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
 CC RX MEDLINE=92084686; PubMed=1748660;  
 CC RA Kang C.-H., Shin W.-C., Yamagata Y., Gokcen S., Ames G.F.-L., Kim S.-H.;

RT "Crystal structure of the lysine-, arginine-, ornithine-binding protein (Lao) from Salmonella typhimurium at 2.7-A resolution."; J. Biol. Chem. 266:23893-23899(1991).  
 RL [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RX MEDLINE=93266593; PubMed=8496186;  
 RA Oh B.-H., Pandit J., Kang C.-H., Nikaido K., Gokcen S., Ames G.F.-L., Kim S.-H.;  
 RA "Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand."; J. Biol. Chem. 268:11348-11355(1993).  
 RL [5]  
 RP ERATUM.  
 RA Oh B.-H., Pandit J., Kang C.-H., Nikaido K., Gokcen S., Ames G.F.-L., Kim S.-H.;  
 RL J. Biol. Chem. 268:17648-17649(1993).  
 CC -1- FUNCTION: THIS PERIPLASMIC BINDING PROTEIN IS INVOLVED IN AN ARGININE TRANSPORT SYSTEM. ARG1 AND HISTIDINE-BINDING PROTEIN J (HISJ) INTERACT WITH A COMMON MEMBRANE-BOUND RECEPTOR, HIGP.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- SIMILARITY: Belongs to the bacterial extracellular solute-binding protein family 3.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; V01368; CAA24651.1; -.  
 CC EMBL; J01805; AAA75577.1; -.  
 CC EMBL; AE008806; AAL21256.1; -.  
 CC PIR; A03408; JKBT.  
 CC PDB; 1LST; 22-JUN-94.  
 CC PDB; 2LQO; 22-JUN-94.  
 CC PDB; 1LAF; 10-JUL-95.  
 CC PDB; 1LAG; 10-JUL-95.  
 CC PDB; 1LAH; 10-JUL-95.  
 CC StyGene; SGI0018; argT.  
 CC InterPro; IPR005768; Lys\_Arg\_Orn\_bind.  
 CC InterPro; IPR001311; SBP/Glu\_receptor.  
 CC InterPro; IPR001638; SBP\_bac3.  
 CC Pfam; PF00497; SBP\_bac3; 1.  
 CC SMART; SM00062; PBpb; 1.  
 CC TIGRFAMs; TIGR01096; 3A010303R; 1.  
 CC PROSITE; PS01039; SBP\_BACTERIAL\_3; 1.  
 CC Transport; Amino-acid transport; Periplasmic; Signal; 3D-structure;  
 CC Complete proteome.  
 CC SIGNAL 1 22  
 CC CHAIN 23 260  
 CC  
 CC DISULFID 60 67  
 CC CONFLICT 124 124  
 CC STRAND 27 32  
 CC STRAND 36 36  
 CC TURN 37 39  
 CC STRAND 40 42  
 CC TURN 44 45  
 CC STRAND 48 49  
 CC HELIX 51 63  
 CC TURN 64 64  
 CC STRAND 66 71  
 CC HELIX 74 76  
 CC TURN 77 82  
 CC TURN 83 84  
 CC STRAND 88 89  
 CC STRAND 94 94  
 CC HELIX 97 102  
 CC STRAND 104 105  
 CC STRAND 110 110  
 CC  
 CC LYSINE-ARGININE-ORNITHINE-BINDING PERIPLASMIC PROTEIN.  
 CC I -> V (IN REF. 1).

```

FT STRAND 114 119
FT TURN 120 121
FT HELIX 128 131
FT TURN 132 133
FT STRAND 135 139
FT TURN 140 141
FT HELIX 143 151
FT TURN 152 152
FT HELIX 153 155
FT TURN 156 156
FT STRAND 158 162
FT HELIX 165 173
FT TURN 174 175
FT STRAND 179 183
FT HELIX 184 190
FT TURN 191 193
FT HELIX 195 197
FT TURN 198 199
FT STRAND 200 202
FT STRAND 207 207
FT HELIX 210 213
FT STRAND 217 217
FT STRAND 220 221
FT TURN 223 224
FT HELIX 226 241
FT TURN 242 243
FT HELIX 244 250
FT TURN 251 252
FT TURN 257 258
SQ SEQUENCE 260 AA; 28200 MW; B996DE523F5F80A5 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 260;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 46 GEFIGFDI 53

RESULT 53
PRTB SCVLI STANDARD; PRT; 260 AA.
AC P15369; Q92333; Q9P962;
DT 01-APR-1990 (Rel. 14, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Scytalidopepsin B precursor (EC 3.4.23.32) (Acid protease B) (SLB).
OS Scytalidium lignicolum.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Scytalidium.
OX NCBI_TaxID=5539;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98430139; PubMed=9757573;
RA Oda N., Gotoh Y., Oyama H., Murao S., Oda K., Tsuru D.;
RT "Nucleotide sequence of the gene encoding the precursor protein of
peptstatin-insensitive acid protease B, scytalidopepsin B, from
Scytalidium lignicolum."
RL Biosci. Biotechnol. Biochem. 62:1637-1639(1998).
RN [2]
RP SEQUENCE OF 50-260 FROM N.A.
RX MEDLINE=96376187; PubMed=8782420;
RA Kakimori T., Yoshimoto T., Oyama H., Oda N., Gotoh Y., Oda K.,
RA Murao S., Tsuru D.;
RT "Nucleotide sequence of the gene encoding pepstatin-insensitive acid
protease B, Scytalidopepsin B, of Scytalidium lignicolum."
RL Biosci. Biotechnol. Biochem. 60:1210-1211(1996).
RN [3]
RP SEQUENCE OF 55-260.
RX MEDLINE=84185536; PubMed=6370989;
RA Maifu T., Nagata S., Matsuda G., Maruta S., Oda K., Murao S.,
RA Tsuru D.;
RT "Complete amino acid sequence of Scytalidium lignicolum acid protease

```

---

```

RT B.";
RL J. Biochem. 95:465-475(1994).
CC -|- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity,
cleaving 24-Phe-|-Phe-25, but not 15-Leu-|-Tyr-16 and 25-Phe-|-
Tyr-26 in the B chain of insulin.
CC -|- SUBUNIT: Monomer.
CC -|- SIMILARITY: Belongs to peptidase family A4.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AB038553; BAA92164.1; -.
CC EMBL; D83963; BAA12157.1; -.
CC PIR; A28864; A28864.
CC PIR; JC4883; JC4883.
CC PIR; JE0300; JE0300.
CC MEROPS; A04.001.1.20
CC InterPro; IPR000250; Peptidase A4.
CC Pfam; PF01828; Peptidase A4; 1.
CC PRINTS; PR00977; SCYTLDPASE.
CC ProDom; PD018627; AsproteaseA4; 1.
CC Hydrolase; Aspartyl protease; Zymogen; Signal.
FT SIGNAL 1 20
FT PROPEP 21 54
FT CHAIN 55 260 SCYTALIDOPEPSIN B.
FT ACT SITE 107 107 POTENTIAL.
FT ACT SITE 152 152 POTENTIAL.
FT DISULFID 101 181
FT DISULFID 195 219
FT DISULFID 248 257
FT CONFLICT 83 91
FT CONFLICT 179 179
FT CONFLICT 198 198
FT CONFLICT 202 202 MISSING (IN REF. 3).
SQ SEQUENCE 260 AA; 27165 MW; D73852833694C6E0 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 260;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 185 AEFITDF 192

RESULT 54
FHUF_ECOLI STANDARD; PRT; 262 AA.
ID FHUF_ECOLI
AC P39405;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ferric iron reductase protein fhuf.
GN FHUF OR B4367.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.I.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes."
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]

```



```

RP CHARACTERIZATION.
RX MEDLINE=99144129; PubMed=9990318;
RA Muller K., Matzanke B.F., Schunemann V., Trautwein A.X., Hantke K.;
RT "FnuF", an iron-regulated protein of Escherichia coli with a new type
RT of 2Fe-2S center."
RL Eur. J. Biochem. 258:1001-1008(1998).
CC -!- FUNCTION: Involved in the reduction of ferric iron in cytoplasmic
CC ferrioxamine B.
CC -!- COFACTOR: Binds 1 2Fe-2S cluster.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; sometime membrane-associated.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/ or
CC send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14003; AAA97266.1; -.
DR EMBL; AE000507; AAC77323.1; -.
DR PIR; S56594; S56594.
DR EcoGene; EG12595; fhuF.
DR InterPro; IPR008090; Re iron reduct.
DR PRINTS; PR01714; 2FE2SRDCTASE.
KW Metal-binding; Iron-sulfur; Iron; 2Fe-2S; Complete proteome.
FT METAL 244 244 IRON-SULFUR (2FE-2S).
FT METAL 245 245 IRON-SULFUR (2FE-2S).
FT METAL 256 256 IRON-SULFUR (2FE-2S).
FT METAL 259 259 IRON-SULFUR (2FE-2S).
SQ SEQUENCE 262 AA; 30112 MW; 7708E776B1BE7793 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 262;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 41 LEFIRLDE 48

RESULT 55
RS4 ARATH
ID RS4 ARATH STANDARD; PRT; 262 AA.
AC P49204; Q42346; Q9FL39; Q9LYQ6;
DT 01-FEB-1996 (Rel. 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 40S ribosomal protein S4.
GN RPS4 OR ATSG07090 OR MON9.26 OR T28J14 30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=98344145; PubMed=9679202;
RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
RT features of the regions of 1,361,565 bp covered by twenty one
RT physically assigned P1 and TAC clones."
RL DNA Res. 5:131-145(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,

```

```

RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang B., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strömatt C.,
RA Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedhia N.,
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,
RA Martensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entian K.-D., Terryn R., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McCullagh B., Robben J., Grymoprez B., Zimmermann W.,
RA Ransberger U., Wedler H., Balke K., Wedler E., Peters S.,
RA van Staveren M., Dirkse W., Moolijman P., Klein Lankhorst R.,
RA Weitzenecker T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,
RA Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransch P.F.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT thaliana."
RL Nature 408:823-826(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J., Cheuk R.F.,
RA Southwick A.M., Wu H.-C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.T., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome."
RL Science 302:842-846(2003).
RN [4]
RP SEQUENCE OF 1-122 FROM N.A.
RC STRAIN=cv. Columbia;
RA Philipps G., Gigot C.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-72 AND 198-262 FROM N.A.
RC STRAIN=cv. Columbia;
RA Raynal M., Grellet F., Laudie M., Meyer Y., Cooke R., Delseny M.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 133-261 FROM N.A.
RC STRAIN=cv. Columbia;
RA Cooke R., Laudie M., Raynal M., Delseny M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 236-262 FROM N.A.
RC STRAIN=cv. Columbia;
RA Berthomieu P., Guerrier D., Giraudat J.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the S4E family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; AB010697; BAB11167.1; ALT\_SEQ.  
CC EMBL; AL163652; CAB87265.1; -.  
CC EMBL; AY050933; AAK93610.1; -.  
CC EMBL; Z17994; CAA79086.1; ALT\_INIT.  
CC EMBL; Z32619; CAA83563.1; -.  
CC EMBL; Z32618; CAA83562.1; -.  
CC EMBL; F20029; CAA23384.1; -.  
CC EMBL; Z18498; CAA79206.1; -.  
CC PIR; T48480; T48480.  
CC InterPro; IPR005824; KOW.  
CC InterPro; IPR006646; KOW\_sub.  
CC InterPro; IPR000876; Ribosomal\_S4E.  
CC InterPro; IPR002942; S4.  
CC Pfam; PF00467; KOW; 1.  
CC Pfam; PF00900; Ribosomal\_S4E; 1.  
CC Pfam; PF01479; S4; 1.  
CC ProDom; PD002667; Ribosomal\_S4E; 1.  
CC SMART; SMO0739; KOW; 1.  
CC SMART; SMO0363; S4; 1.  
CC PROSITE; PS00528; RIBOSOMAL\_S4E; 1.  
CC PROSITE; PS00889; S4; 1.  
KW Ribosomal protein; RNA-binding.  
FT DOMAIN 42 104 S4 RNA-BINDING.  
FT CONFLICT 28 28 A -> G (IN REF. 5).  
FT CONFLICT 31 31 P -> S (IN REF. 5).  
FT CONFLICT 53 53 K -> M (IN REF. 5).  
FT CONFLICT 55 55 A -> S (IN REF. 5).  
FT CONFLICT 92 92 I -> IV (IN REF. 4).  
FT CONFLICT 257 258 AS -> SA (IN REF. 6).  
SQ SEQUENCE 262 AA; 29874 MW; EF0847171FE9CA67 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 262;  
Best Local Similarity 50.0%; Pred. No. 2.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
Db 170 VEFIKFDV 177

RESULT 56  
RNC\_BACHD  
ID RNC\_BACHD STANDARD; PRT; 263 AA.  
AC Q9KA05;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Ribonuclease III (EC 3.1.26.3) (RNase III).  
GN RNC OR RNCB OR BH2489.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirana C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
CC -!- FUNCTION: Digests double-stranded RNA. Involved in the processing  
CC of ribosomal RNA precursors and of some mRNAs (By similarity).  
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-  
CC phosphomonoester.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- SIMILARITY: Contains 1 DRBM (double-stranded RNA-binding) domain.  
CC -!- SIMILARITY: Contains 1 RNase III domain.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; AP001515; BAB06208.1; -.  
CC PIR; A83961; A83961.  
CC HAMAP; MF\_00104; -; 1.  
CC InterPro; IPR001159; DS\_RBD.  
CC InterPro; IPR000999; RNase\_III.  
CC Pfam; PF00035; dsrm; 1.  
CC Pfam; PF00636; Ribonuclease\_3; 1.  
CC SMART; SMO0358; DSRM; 1.  
CC SMART; SMO0535; RIBOC; 1.  
CC PROSITE; PS00137; DS\_RBD; 1.  
CC PROSITE; PS00517; RNase\_3\_1; 1.  
CC PROSITE; PS00142; RNase\_3\_2; 1.  
KW Hydrolase; Nuclease; Endonuclease; RNA-binding; Complete proteome.  
FT DOMAIN 35 164 RNase III.  
FT DOMAIN 190 259 DRBM.  
SQ SEQUENCE 263 AA; 30224 MW; 61B8E1C557CC5485 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 263;  
Best Local Similarity 50.0%; Pred. No. 2.9e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8  
Db 196 QEFIQRDN 203

RESULT 57  
MIND\_ECOLI  
ID MIND\_ECOLI STANDARD; PRT; 269 AA.  
AC P18197;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Septum site-determining protein mind (Cell division-inhibitor mind).  
GN MIND OR B1175 OR C1622 OR Z1937 OR ECS1669 OR SF1162 OR S1248.  
OS Escherichia coli.  
OS Escherichia coli O6.  
OS Escherichia coli O157:H7, and  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562, 217992, 83334, 623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli;  
RX MEDLINE=89136010; PubMed=2645057;  
RA de Boer P.A.J., Crossley R.E., Rothfield L.I.;  
RT "A division inhibitor and a topological specificity factor coded for  
RT by the minicell locus determine proper placement of the division  
RT septum in E. coli.";  
RL Cell 56:641-649(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;

RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.,  
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
[4]  
RN SEQUENCE FROM N.A.  
RP SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;  
RX MEDLINE=2238234; PubMed=12471157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
RA Mobley G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
RT "Extensive mosaic structure revealed by the complete genome sequence  
of uropathogenic *Escherichia coli*.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
[5]  
RN SEQUENCE FROM N.A.  
RP SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";  
RL Nature 409:529-533(2001).  
[6]  
RN SEQUENCE FROM N.A.  
RP SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Tida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;  
RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
[7]  
RN SEQUENCE OF 1-12.  
RP SPECIES=E.coli; STRAIN=K12 / EMG2;  
RX MEDLINE=97443975; PubMed=92298645;  
RA Link A.J., Robison K., Church G.M.;  
RT "Comparing the predicted and observed properties of proteins encoded  
in the genome of *Escherichia coli* K-12.";  
RL Electrophoresis 18:1259-1313(1997).  
[8]  
RN SEQUENCE FROM N.A.  
RP SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
RX MEDLINE=22272406; PubMed=12384590;  
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.;  
RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
through comparison with genomes of *Escherichia coli* K12 and O157.";  
RL Nucleic Acids Res. 30:4432-4441(2002).  
[9]  
RN SEQUENCE FROM N.A.  
RP SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
RX MEDLINE=22590274; PubMed=12704152;  
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
RA Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of *Shigella*  
*flexneri* serotype 2a strain 2457T.";

RL Infect. Immun. 71:2775-2786(2003).  
RN [10]  
RP FUNCTION, AND MUTAGENESIS.  
RX SPECIES=E.coli;  
RA MEDLINE=92097557; PubMed=1836760;  
RA de Boer P.A.J., Crossley R.E., Hand A.R., Rothfield L.I.;  
RT "The Mind protein is a membrane ATPase required for the correct  
placement of the *Escherichia coli* division site.";  
RL EMBO J. 10:4371-4380(1991).  
[11]  
RN CHARACTERIZATION.  
RP SPECIES=E.coli;  
RX MEDLINE=92384467; PubMed=10220403;  
RA Raskin D.M., de Boer P.A.;  
RT "Rapid pole-to-pole oscillation of a protein required for directing  
division to the middle of *Escherichia coli*.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:4971-4976(1999).  
CC -!- FUNCTION: ATPASE REQUIRED FOR THE CORRECT PLACEMENT OF THE  
DIVISION SITE. CELL DIVISION INHIBITORS MINC AND MIND ACT IN  
CONCERT TO FORM AN INHIBITOR CAPABLE OF BLOCKING FORMATION OF THE  
POLAR Z RING SEPTUMS. RAPIDLY OSCILLATES BETWEEN THE POLES OF THE  
CELL TO DESTABILIZE FTSZ FILAMENTS THAT HAVE FORMED BEFORE THEY  
MATURE INTO POLAR Z RINGS.  
CC -!- SUBUNIT: Interacts with minC and ftsZ.  
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated.  
CC -!- SIMILARITY: Belongs to the para family. Mind subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J03153; AAB59062.1; -;  
DR EMBL; A5000216; AAC74259.1; -;  
DR EMBL; D90751; BAA36009.1; -;  
DR EMBL; D90752; BAA36022.1; -;  
DR EMBL; A5016759; AAN80087.1; -;  
DR EMBL; A5005334; AAG56026.1; -;  
DR EMBL; A5003555; BAA35092.1; -;  
DR EMBL; A5015142; AAN42778.1; -;  
DR EMBL; A5016982; AAP16670.1; -;  
DR PIR; B31877; CCECID.  
DR PIR; F85695; F85695.  
DR PIR; F85695; F85695.  
DR SWISS-2DPAGE; P18197; COLI.  
DR EcoGene; E310597; mind.  
DR InterPro; IPR003348; Arsa\_ATPase.  
DR InterPro; IPR000707; ATPase\_ParA.  
DR Pfam; PF02374; Arsa\_ATPase; 1.  
DR Pfam; PF00991; ParA; 1.  
KW Cell division; Septation; ATP-binding; Inner membrane;  
KW Complete proteome.  
FT INIT MET 0  
FT NP BIND 9 16 ATP (POTENTIAL).  
FT MUTAGEN 14 14 G->S: LESS EFFECTIVE THEN WILD-TYPE.  
FT MUTAGEN 15 15 K->Q: LOSS OF ACTIVITY.  
FT MUTAGEN 15 16 KT->QR: LOSS OF ACTIVITY.  
SQ SEQUENCE 269 AA; 29483 MW; 0D1B29A476A190B1 CRC64;  
Query Match 100.0%; Score 25; DB 1; Length 269;  
Best Local Similarity 50.0%; Pred. No. 3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 1 XEFIXYX 8  
DB 113 FEIVCDS 120  
RESULT 58  
ILIA\_RAT

ID IL1A RAT STANDARD; PRT; 270 AA.  
 AC F16598;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Interleukin-1 alpha precursor (IL-1 alpha).  
 GN IL1A.  
 OS Rattus norvegicus (Rat).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89278054; PubMed=2471704;  
 RA Nishida T., Nishino N., Takano M., Sekiguchi Y., Kawai K.,  
 RA Mizuno K., Nakai S., Masui Y., Hirai Y.;  
 RT "Molecular cloning and expression of rat interleukin-1 alpha cDNA.";  
 RL J. Biochem. 105:351-357 (1989).  
 CC -!- FUNCTION: PRODUCED BY ACTIVATED MACROPHAGES, IL-1 STIMULATES  
 CC THYMOCYTE PROLIFERATION BY INDUCING IL-2 RELEASE, B-CELL  
 CC MATURATION & PROLIFERATION, & FIBROBLAST GROWTH FACTOR ACTIVITY.  
 CC IL-1 PROTEINS ARE INVOLVED IN THE INFLAMMATORY RESPONSE, BEING  
 CC IDENTIFIED AS ENDOGENOUS PYROGENS, AND ARE REPORTED TO STIMULATE  
 CC THE RELEASE OF PROSTAGLANDIN AND COLLAGENASE FROM SYNOVIAL CELLS.  
 CC -!- SUBUNIT: MONOMER.  
 CC -!- DOMAIN: THE SIMILARITY AMONG THE IL-1 PRECURSORS SUGGESTS THAT THE  
 CC AMINO ENDS OF THESE PROTEINS SERVE SOME AS YET UNDEFINED FUNCTION.  
 CC -!- MISCELLANEOUS: THE LACK OF A SPECIFIC HYDROPHOBIC SEGMENT IN THE  
 CC PRECURSOR SEQUENCE SUGGESTS THAT IL-1 IS RELEASED BY DAMAGED CELLS  
 CC OR IS SECRETED BY A MECHANISM DIFFERING FROM THAT USED FOR OTHER  
 CC SECRETORY PROTEINS.  
 CC -!- SIMILARITY: Belongs to the IL-1 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D00403; BAA00306.1; .  
 CC PIR: JX0064; JX0064.  
 CC HSP: F01583; I17A.  
 CC InterPro: IPR008996; Cytok\_IL1\_like.  
 CC InterPro: IPR003502; IL1\_propep.  
 CC InterPro: IPR000975; Interleukin\_1.  
 CC Pfam: PF00340; IL1; 1.  
 CC Pfam: PF02394; IL1\_propep; 1.  
 CC SMART: SM00125; IL1; 1.  
 CC PROSITE: PS00253; INTERLEUKIN\_1; 1.  
 CC Cytokine; Macrophage; Mitogen; Inflammatory response; Pyrogen.  
 FT PROPEP 1 114  
 FT CHAIN 115 270 INTERLEUKIN-1 ALPHA.  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 270 AA; 30856 MW; E3EB22F57F291DE CRC64;  
 Query Match 100.0%; Score 25; DB 1; Length 270;  
 Best Local Similarity 50.0%; Pred. No. 3e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEPIXXDX 8  
 :|||||:  
 Db 134 QEFIMNDS 141  
 RESULT 59  
 IL1A HUMAN  
 ID IL1A HUMAN STANDARD; PRT; 271 AA.  
 AC P01583;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Interleukin-1 alpha precursor (IL-1 alpha) (Hematopoietin-1).  
 GN IL1A.  
 OS Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85240547; PubMed=2989698;  
 RA March C.J., Mosley B., Larsen A., Cerrretti D.P., Braedt G., Price V.,  
 RA Gillis S., Henney C.S., Kronhelm S.R., Grabstein K., Conlon P.J.,  
 RA Hopp T.P., Cosman D.;  
 RT "Cloning, sequence and expression of two distinct human interleukin-1  
 RT complementary DNAs.";  
 RL Nature 315:641-647 (1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86205226; PubMed=3486405;  
 RA Furutani Y., Notake M., Fukui T., Ohue M., Nomura H., Yamada M.,  
 RA Nakamura S.;  
 RT "Complete nucleotide sequence of the gene for human interleukin 1  
 RT alpha.";  
 RL Nucleic Acids Res. 14:3167-3179 (1986).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND VARIANT SER-114.  
 RX MEDLINE=85297782; PubMed=2994016;  
 RA Furutani Y., Notake M., Yamayoshi M., Yamagishi J., Nomura H.,  
 RA Ohue M., Furuta R., Fukui T., Yamada M., Nakamura S.;  
 RT "Cloning and characterization of the cDNAs for human and rabbit  
 RT interleukin-1 precursor.";  
 RL Nucleic Acids Res. 13:5869-5882 (1985).  
 RN [4]  
 RP SEQUENCE FROM N.A., AND VARIANT SER-114.  
 RX MEDLINE=90249285; PubMed=2635664;  
 RA Kofenko S.V., Buleikov M.T., Veiko V.P., Epishin S.M., Lomakin I.B.,  
 RA Emel'yanov A.V., Kozlov A.P., Konusova V.G., Kotov A.Y.,  
 RA Kurbatova T.V., Reshetnikov V.I., Simbirtsev A.S., Katlinskii S.A.,  
 RA Vinetskii Y.P.;  
 RT "Cloning of the cDNA coding for human prointerleukin-1 alpha and  
 RT prointerleukin-1 beta.";  
 RL Dokl. Akad. Nauk SSSR 309:1005-1008 (1989).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86141855; PubMed=3485152;  
 RA Gubler U., Chua A.O., Stern A.S., Hellmann C.P., Vitek M.P.,  
 RA Dechiara T.M., Benjamin W.R., Collier K.J., Dukovich M.,  
 RA Familletti P.C., Fiedler-Nagy C., Jensen J., Kaffka K., Kilian P.L.,  
 RA Stremlo D., Wittreich B.H., Woehle D., Mizel S.B., Lomedico P.T.;  
 RT "Recombinant human interleukin 1 alpha: purification and biological  
 RT characterization.";  
 RL J. Immunol. 136:2492-2497 (1986).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87156769; PubMed=3493774;  
 RA Nishida T., Nishino N., Takano M., Kawai K., Masui Y.,  
 RA Nakai S., Hirai Y.;  
 RT "cDNA cloning of IL-1 alpha and IL-1 beta from mRNA of U937 cell  
 RT line.";  
 RL Biochem. Biophys. Res. Commun. 143:345-352 (1987).  
 RN [7]  
 RP SEQUENCE FROM N.A., AND VARIANTS GLN-85; SER-114 AND ASN-138.  
 RA Rieder M.J., Armet T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,  
 RA Rajkumar N.R., Toth E.J., Yi Q., Nickerson D.A.;  
 RL Submitted (AUG-2002) to the EMBL/Genbank/DBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Villard J., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Faley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Blakesley R.C., Griwood J., Schmutz J., Myers R.M.,  
RA Buterfield V.S.N., Krzyzinski M.I., Skalska U., Smalhus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RA "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RP SEQUENCE OF 113-132.  
RX MEDLINE=89184226; PubMed=3281727;  
RA Zsebo K.M., Wypych J., Yushchenko V.N., Lu H., Hunt P., Dukes P.P.,  
RA Langley K.E.;  
RT "Effects of hematopoietin-1 and interleukin 1 activities on early  
RT hematopoietic cells of the bone marrow.";  
RL Blood 71:962-968 (1988).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RX MEDLINE=90268016; PubMed=2346741;  
RA Graves B.J., Hatada M.H., Hendrickson W.A., Miller J.K., Madison V.S.,  
RA Satow Y.;  
RT "Structure of interleukin 1 alpha at 2.7-A resolution.";  
RL Biochemistry 29:2679-2684 (1990).  
RN [11]  
RP MYRISTOYLATION.  
RX MEDLINE=93348250; PubMed=8346241;  
RA Stevenson F.T., Bursten S.L., Fanton C., Locksley R.M., Lovett D.H.;  
RT "The 31-kDa precursor of interleukin 1 alpha is myristoylated on  
RT specific lysines within the 16-kDa N-terminal piece.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:7245-7249 (1993).  
CC -!- FUNCTION: PRODUCED BY ACTIVATED MACROPHAGES. IL-1 STIMULATES  
CC THYMOCYTE PROLIFERATION BY INDUCING IL-2 RELEASE, B-CELL  
CC MATURATION & PROLIFERATION, & FIBROBLAST GROWTH FACTOR ACTIVITY.  
CC IL-1 PROTEINS ARE INVOLVED IN THE INFLAMMATORY RESPONSE, BEING  
CC IDENTIFIED AS ENDOGENOUS PYROGENS, AND ARE REPORTED TO STIMULATE  
CC THE RELEASE OF PROSTAGLANDIN AND COLLAGENASE FROM SYNOVIAL CELLS.  
CC -!- SUBUNIT: Monomer.  
CC -!- DOMAIN: THE SIMILARITY AMONG THE IL-1 PRECURSORS SUGGESTS THAT THE  
CC AMINO ENDS OF THESE PROTEINS SERVE SOME AS YET UNDEFINED FUNCTION.  
CC -!- MISCELLANEOUS: THE LACK OF A SPECIFIC HYDROPHOBIC SEGMENT IN THE  
CC PRECURSOR SEQUENCE SUGGESTS THAT IL-1 IS RELEASED BY DAMAGED CELLS  
CC OR IS SECRETED BY A MECHANISM DIFFERING FROM THAT USED FOR OTHER  
CC SECRETORY PROTEINS.  
CC -!- SIMILARITY: Belongs to the IL-1 family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X02531; CAA26371.1; -  
DR EMBL; X03833; CAA27448.1; -  
DR EMBL; X02851; CAA26604.1; -  
DR EMBL; X56086; CAA39566.1; -  
DR EMBL; M28983; AAA59134.1; -  
DR EMBL; M15329; AAA59133.1; -  
DR EMBL; AF536338; AAM96189.1; -  
DR EMBL; BC013142; AAH13142.1; -  
DR PIR; A23385; ICHUIA.  
DR PDB; 2ILA; 15-OCT-92.  
DR PDB; 1ITA; 31-OCT-93.  
DR Genew; HGNC:5991; IL1A.

DR MIM; 147760; -; C:extracellular space; TAS.  
DR GO; GO:0005615; F:cytokine activity; TAS.  
DR GO; GO:0004871; F:signal transducer activity; TAS.  
DR GO; GO:0006916; P:anti-apoptosis; TAS.  
DR GO; GO:0008283; P:cell proliferation; TAS.  
DR GO; GO:0006954; P:inflammatory response; TAS.  
DR GO; GO:0008285; P:negative regulation of cell proliferation; TAS.  
DR InterPro; IPR008996; Cytok IL1 like.  
DR InterPro; IPR003502; IL1\_propep.  
DR InterPro; IPR000975; Interleukin\_1.  
DR Pfam; PF00340; IL1; 1.  
DR Pfam; PF02394; IL1\_propep; 1.  
DR SMART; SM00125; IL1; 1.  
DR PROSITE; PS00253; INTERLEUKIN\_1; 1.  
KW Cytokine; Macrophage; Mitogen; Inflammatory response; Pyrogen;  
KW Polymorphism; 3D-structure; Myristate; Lipoprotein.  
FT PROPEP 1 112 INTERLEUKIN-1 ALPHA.  
FT CHAIN 113 271 N(6)-myristoyl lysine.  
FT LIPID 82 83 N(6)-myristoyl lysine.  
FT CARBOHYD 102 102 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 141 141 N-LINKED (GLCNAC...) (POTENTIAL).  
FT VARIANT 85 85 R->Q.  
FT /FTID=VAR\_014304.  
FT VARIANT 114 114 A->S (in dbSNP:17561).  
FT /FTID=VAR\_014305.  
FT VARIANT 125 125 N->D (in dbSNP:17562).  
FT /FTID=VAR\_014600.  
FT VARIANT 138 138 D->N.  
FT /FTID=VAR\_014306.  
FT VARIANT 176 176 D->H (in dbSNP:1801715).  
FT /FTID=VAR\_014601.  
FT STRAND 124 138  
FT STRAND 143 146  
FT STRAND 152 155  
FT TURN 161 162  
FT STRAND 166 172  
FT STRAND 181 186  
FT STRAND 191 194  
FT STRAND 203 206  
FT STRAND 213 214  
FT HELIX 216 221  
FT STRAND 222 228  
FT TURN 229 230  
FT STRAND 231 236  
FT TURN 240 241  
FT STRAND 243 244  
FT STRAND 251 251  
FT STRAND 254 255  
Query Match 100.0%; Score 25; DB 1; Length 271;  
Best Local Similarity 50.0%; Pred No. 3e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 4; Conservative 4;  
QY 1 XEFIXDX 8  
:||||:  
Db 132 YEFILNDA 139  
RESULT 60  
IL1A MACFA  
ID IL1A MACFA STANDARD; PRT; 271 AA.  
AC P79340;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Interleukin-1 alpha precursor (IL-1 alpha) (Hematopoietin-1).  
GN IL1A.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Macaca.

NCBI\_TaxID=9541; [1]

RNN  
RP  
RRL

Tatsumi M.;  
Submitted RRA

- FUNCTION: PRODUCED BY ACTIVATED MACROPHAGES, IL-1 STIMULATES THYMOCYTE PROLIFERATION BY INDUCING IL-2 RELEASE, B-CELL MATURATION & PROLIFERATION, & FIBROBLAST GROWTH FACTOR ACTIVITY.  
IL-1 PROTEINS ARE INVOLVED IN THE INFLAMMATORY RESPONSE, BEING IDENTIFIED AS ENDOGENOUS PYROGENS, AND ARE REPORTED TO STIMULATE THE RELEASE OF PROSTAGLANDIN AND COLLAGENASE FROM SYNOVIAL CELLS.

-- SUBUNIT: Monomer.  
-- DOMAIN: THE SIMILARITY AMONG THE IL-1 PRECURSORS SUGGESTS THAT THE AMINO ENDS OF THESE PROTEINS SERVE SOME AS YET UNDEFINED FUNCTION.

-- MISCELLANEOUS: THE LACK OF A SPECIFIC HYDROPHOBIC SEGMENT IN THE PRECURSOR SEQUENCE SUGGESTS THAT IL-1 IS RELEASED BY DAMAGED CELLS OR IS SECRETED BY A MECHANISM DIFFERING FROM THAT USED FOR OTHER SECRETORY PROTEINS.

-- SIMILARITY: Belongs to the IL-1 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpost at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licenses@isb-sib.ch).

EMBL; AB000553; BAAL9147.1; -;  
HSSP; P01583; IITA.  
InterPro: IPR008996; Cytok\_IL1\_like.  
InterPro: IPR003502; IL1\_propep.  
InterPro: IPR000975; Interleukin\_1.  
Pfam; PF003340; IL1; 1.  
Pfam; PF023394; IL1\_propep; 1.  
SMART; SMW0125; ILI; 1.  
DR DR PROSITE; PS00253; INTERLEUKIN\_1; 1.  
KW Cytokine; Macrophage; Mitogen; Inflammatory response; Pyrogen.  
FT PROPEP 1 112 BY SIMILARITY.  
FT CHAIN 113 271 INTERLEUKIN-1 ALPHA.  
FT CARBOHYD 102 102 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 121 121 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 137 137 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 141 141 N-LINKED (GLCNAC... ) (POTENTIAL).  
SEQUENCE 271 AA; 03685 MW; 0526A9A6404558F0 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 271;  
Best Local Similarity 50.0%; Pred. No. 3e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0

Oy 1 KEFIIXDX 8  
          |||||:  
Db 132 HEFILNDT 139

RESULT 61

IL1A MACMU  
ID ID IL1A MACMU STANDARD; PRT; 271 AA.  
AC AC P48089;  
DT DT 01-FEB-1996 (Rel. 33, Created)  
DI DI 01-FEB-1996 (Rel. 33, Last sequence update)  
DT DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DE Interleukin-1 alpha precursor (IL-1 alpha) (Hematopoietin-1).  
GN IL1A.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;  
CC Cercopitheciinae; Macaca.  
OX NCBI\_TaxID=9544;  
RN RN [1]  
RS  
RP SEQUENCE FROM N.A.  
RC TISSUE=Blood;  
RX MEDLINE=96003435; PubMed=7561102;

CC -!- FUNCTION: THIS IS A COPPER-CONTAINING OXIDASE THAT FUNCTIONS IN  
 CC THE FORMATION OF PIGMENTS SUCH AS MELANINS AND OTHER POLYPHENOLIC  
 CC COMPOUNDS.  
 CC -!- CATALYTIC ACTIVITY: L-tyrosine + L-DOPA + O(2) = L-DOPA +  
 CC DOPAQINONE + H(2)O.  
 CC -!- COFACTOR: Binds 2 copper ions per subunit (By similarity).  
 CC -!- SIMILARITY: Belongs to the tyrosinase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X95703; CAA65000.1; -;  
 CC InterPro; IPR008922; Di-copper\_centre.  
 CC InterPro; IPR002227; Tyrosinase.  
 CC Pfam; PF00264; tyrosinase; 1.  
 CC PRINTS; PR00092; TYROSINASE.  
 CC PROSITE; PS00497; TYROSINASE\_1; 1.  
 CC PROSITE; PS00498; TYROSINASE\_2; 1.  
 CC Melanin biosynthesis; Oxidoreductase; Monooxygenase; Copper.  
 CC INIT MET 0  
 CC METAL 36 36 COPPER A (BY SIMILARITY).  
 CC METAL 52 52 COPPER A (BY SIMILARITY).  
 CC METAL 61 61 COPPER A (BY SIMILARITY).  
 CC METAL 188 188 COPPER B (BY SIMILARITY).  
 CC METAL 192 192 COPPER B (BY SIMILARITY).  
 CC METAL 214 214 COPPER B (BY SIMILARITY).  
 CC SEQUENCE 272 AA; 30723 MW; F7F0589CAEB8E0B9 CRC64;  
 CC  
 CC Query Match 100.0%; Score 25; DB 1; Length 272;  
 CC Best Local Similarity 50.0%; Pred. No. 3e+02; Indels 0; Gaps 0;  
 CC Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 XEFIXDX 8  
 CC Db 37 NEFIVADT 44  
 CC  
 CC RESULT 63  
 CC NADE\_STABP STANDARD; PRT; 275 AA.  
 CC AC QBCNF1;  
 CC DT 15-MAR-2004 (Rel. 43, Created)  
 CC DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 CC DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 CC DE NH(3)-dependent NAD(+) synthetase [EC 6.3.5.1].  
 CC GN NADE OR SE1596.  
 CC OS Staphylococcus epidermidis.  
 CC OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 CC OX NCBI\_TaxID=1282;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=ATCC 12228;  
 CC RX PubMed=12950922;  
 CC RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,  
 CC Qian Z.-O., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,  
 CC Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;  
 CC "Genome-based analysis of virulence genes in a non-biofilm-forming  
 CC Staphylococcus epidermidis strain (ATCC 12228).";  
 CC Mol. Microbiol. 49:1577-1593(2003).  
 CC CC -!- CATALYTIC ACTIVITY: ATP + deamido-NAD(+) + NH(3) + H(2)O = AMP +  
 CC diphosphate + NAD(+).  
 CC -!- PATHWAY: NAD biosynthesis.  
 CC -!- SIMILARITY: Belongs to the NAD synthetase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE016749; AAO05195.1; -;  
 CC HAMAP; MF 00193; -; 1.  
 CC InterPro; IPR003694; NAD\_synthase.  
 CC Pfam; PF02540; NAD\_Synthase; 1.  
 CC TIGRFAMs; TIGR00552; nade; 1.  
 CC Ligase; NAD; ATP-binding; Complete proteome.  
 CC NP BIND 47 54 ATP (BY SIMILARITY).  
 CC ACT SITE 49 49 BY SIMILARITY.  
 CC SEQUENCE 275 AA; 30880 MW; CF4B22401F977D9 CRC64;  
 CC  
 CC Query Match 100.0%; Score 25; DB 1; Length 275;  
 CC Best Local Similarity 50.0%; Pred. No. 3.1e+02;  
 CC Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 XEFIXDX 8  
 CC Db 97 LEFINPDT 104  
 CC  
 CC RESULT 64  
 CC TYRO\_STPAL STANDARD; PRT; 275 AA.  
 CC ID TYRO\_STPAL  
 CC AC PS5022;  
 CC DT 01-OCT-1996 (Rel. 34, Created)  
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Tyrosinase [EC 1.14.18.1] (Monophenol monooxygenase).  
 CC GN MELC2 OR MEL.  
 CC OS Streptomyces albus G.  
 CC OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 CC OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 CC OX NCBI\_TaxID=1962;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=DSM 40480;  
 CC RA Wehmeier U.F., Brass N., Roessler C., Piepersberg W.;  
 CC Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: THIS IS A COPPER-CONTAINING OXIDASE THAT FUNCTIONS IN  
 CC THE FORMATION OF PIGMENTS SUCH AS MELANINS AND OTHER POLYPHENOLIC  
 CC COMPOUNDS.  
 CC -!- CATALYTIC ACTIVITY: L-tyrosine + L-DOPA + O(2) = L-DOPA +  
 CC DOPAQINONE + H(2)O.  
 CC -!- COFACTOR: Binds 2 copper ions per subunit (By similarity).  
 CC -!- SIMILARITY: Belongs to the tyrosinase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X95705; CAA65005.1; -;  
 CC InterPro; IPR008922; Di-copper\_centre.  
 CC InterPro; IPR002227; Tyrosinase.  
 CC Pfam; PF00264; tyrosinase; 1.  
 CC PRINTS; PR00092; TYROSINASE.  
 CC PROSITE; PS00497; TYROSINASE\_1; 1.  
 CC PROSITE; PS00498; TYROSINASE\_2; FALSE NEG.  
 CC Melanin biosynthesis; Oxidoreductase; Monooxygenase; Copper.  
 CC INIT MET 0  
 CC METAL 37 37 COPPER A (BY SIMILARITY).  
 CC METAL 55 55 COPPER A (BY SIMILARITY).  
 CC METAL 65 65 COPPER A (BY SIMILARITY).  
 CC METAL 192 192 COPPER B (BY SIMILARITY).  
 CC METAL 196 196 COPPER B (BY SIMILARITY).  
 CC METAL 218 218 COPPER B (BY SIMILARITY).  
 CC SEQUENCE 275 AA; 31179 MW; 93998BAEFB93A173 CRC64;



```

Query Match      100.0%; Score 25; DB 1; Length 275;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 38 NEFIMSDT 45

RESULT 65
NADE CORGL STANDARD; PRT; 277 AA.
AC Q8NMN7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NH(3)-dependent NAD(+) synthetase (EC 6.3.5.1).
GN NADE OR CGL2534.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + deamido-NAD(+) + NH(3) + H(2)O = AMP +
CC diposphate + NAD(+).
CC -!- PATHWAY: NAD biosynthesis.
CC -!- SIMILARITY: Belongs to the NAD synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP005282; BAB99927.1; -.
CC HAMAP; MF_00193; -.
CC InterPro; IPR003694; NAD synthase.
CC Pfam; PF02540; NAD synthase; 1.
CC TIGFAMS; TIGR00552; nade; 1.
CC Ligase; NAD; ATP-binding; Complete proteome.
FT NP_BIND 46 53
FT ACT_SITE 48 48 BY SIMILARITY.
FT ACT_SITE 48 48 BY SIMILARITY.
SQ SEQUENCE 277 AA; 30426 MW; 18C1F62CEA756B35 CRC64;

Query Match      100.0%; Score 25; DB 1; Length 277;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 98 LEFIAPDK 105

RESULT 66
KC2B YEAST STANDARD; PRT; 278 AA.
ID KC2B YEAST
AC P43639; Q9URG5;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Casein kinase II beta chain (CK II).
GN CKB1 OR YGL019W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

```

```

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=95256195; PubMed=7737972;
RA Bidwai A.P., Reed J.C., Glover C.V.C.;
RT "Cloning and disruption of CKB1, the gene encoding the 38-kDa beta
RT subunit of Saccharomyces cerevisiae casein kinase II (CKII). Deletion
RT of CKII regulatory subunits elicits a salt-sensitive phenotype."
RL J. Biol. Chem. 270:10395-10404(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Hebling U., Hofmann B., Delius H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 34-52; 182-196 AND 257-270.
RX MEDLINE=94182950; PubMed=8135547;
RA Bidwai A.P., Reed J.C., Glover C.V.C.;
RT "Casein kinase II of Saccharomyces cerevisiae contains two distinct
RT regulatory subunits, beta and beta'."
RL Arch. Biochem. Biophys. 309:348-355(1994).
CC -!- FUNCTION: Plays a complex role in regulating the basal catalytic
CC activity of the alpha subunit (By similarity).
CC -!- SUBUNIT: Tetramer composed of an alpha chain, an alpha', one beta
CC chain and one beta' chain.
CC -!- PTM: Phosphorylated by alpha chain (By similarity).
CC -!- SIMILARITY: Belongs to the casein kinase 2 beta chain family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U21283; AAA86829.1; -.
CC EMBL; Z72541; CRA96719.1; -.
CC PIR; A56421; A56421.
CC GeneOnline; 141067; -.
CC SGD; S0002987; CKB1.
CC DR GO; GO:0005956; C:protein kinase CK2 complex; IDA.
CC DR GO; GO:0004682; P:protein kinase CK2 activity; IDA.
CC DR GO; GO:0006873; P:cell ion homeostasis; IMP.
CC DR GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. .); IPI.
CC DR GO; GO:0000128; P:flocculation; IMP.
CC DR GO; GO:0000082; P:G1/S transition of mitotic cell cycle; IPI.
CC DR GO; GO:0000086; P:G2/M transition of mitotic cell cycle; IPI.
CC DR GO; GO:0006468; P:protein amino acid phosphorylation; IDA.
CC DR GO; GO:0006356; P:regulation of transcription from Pol I prom. .; IDA.
CC DR GO; GO:0006359; P:regulation of transcription from Pol III pr. .; IDA.
CC DR GO; GO:0006974; P:response to DNA damage stimulus; IDA.
CC DR InterPro; IPR000704; CAS_kinase_II.
CC DR Pfam; PF01214; CK_II_beta; 1.
CC DR PRINTS; PR00472; CASKINASEII.
CC DR ProDom; PD003829; CAS kinase II; 1.
CC DR PROSITE; PS01101; CK2_BETA; 1.
CC Transferase; Serine/threonine-protein kinase; Phosphorylation.
SQ SEQUENCE 278 AA; 32265 MW; 9ACA8D285B5990AF CRC64;

Query Match      100.0%; Score 25; DB 1; Length 278;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 44 TEFIEDDF 51

RESULT 67
TRMB_BACTN STANDARD; PRT; 280 AA.
ID TRMB_BACTN
AC Q8A0X7;

```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Z11874; -; NOT ANNOTATED\_CDS.  
DR EMBL; X70810; CAA50077.1; -.  
DR PIR; S34496; S34496.  
KW Chloroplast; Hypothetical protein.  
SQ SEQUENCE 281 AA; 33009 MW; 0B72FCE5982FE320A CRC64;  
  
Query Match 100.0%; Score 25; DB 1; Length 281;  
Best Local Similarity 50.0%; Pred. No. 3.1e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 XEFIXDX 8  
Db 208 NEFIGVDS 215  
:|||||:  
208 NEFIGVDS 215  
  
RESULT 69  
END4\_WOLSU  
ID END4\_WOLSU STANDARD; FRT; 282 AA.  
AC Q7M8B4;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Probable endonuclease IV (EC 3.1.21.2) (Endodeoxyribonuclease IV).  
GN NFO OR WSI754.  
OS Wolinella succinogenes.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
OC Helicobacteraceae; Wolinella.  
NCBI\_TaxID=844;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSMZ 1740;  
RX MEDLINE=22882897; PubMed=14500908;  
RA Baar C., Epinger M., Raddatz G., Simon J., Lanz C., Klimmek O.,  
RA Nandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B.,  
RA Meyer F., Lederer H., Schuster S.C.;  
RT "Complete genome sequence and analysis of Wolinella succinogenes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695(2003).  
CC -!- FUNCTION: Endonuclease IV plays a role in DNA repair. It cleaves  
CC phosphodiester bonds atapurinic or apyrimidinic sites (AP sites)  
CC to produce new 5' ends that are base-free deoxyribose 5-phosphate  
CC residues. It preferentially attacks modified AP sites created by  
CC bleomycin and neocarzinostatin (By similarity).  
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-  
CC phosphooligonucleotide end-products.  
CC -!- COPACATOR: Binds 3 zinc ions (By similarity).  
CC -!- SIMILARITY: Belongs to the AP endonuclease family 2.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; BX571661; CAE10776.1; -.  
DR HAMAP; MF\_00152; -; 1.  
DR PROSITE; PS00723; AP\_NUCLEASE F2.1; 1.  
DR PROSITE; PS00730; AP\_NUCLEASE F2.2; 1.  
DR PROSITE; PS00731; AP\_NUCLEASE F2.3; FALSE NEG.  
KW Hydrolase; Nuclease; Endonuclease; DNA repair; Metal-binding; Zinc;  
KW Complete proteome.  
METAL 70 70 ZINC 1 (BY SIMILARITY).  
METAL 110 110 ZINC 1 (BY SIMILARITY).  
METAL 146 146 ZINC 1 AND 2 (BY SIMILARITY).  
FT

```
FT METAL 180 180 ZINC 2 (BY SIMILARITY).
FT METAL 183 183 ZINC 3 (BY SIMILARITY).
FT METAL 217 217 ZINC 2 (BY SIMILARITY).
FT METAL 230 230 ZINC 3 (BY SIMILARITY).
FT METAL 232 232 ZINC 3 (BY SIMILARITY).
FT METAL 262 262 ZINC 2 (BY SIMILARITY).
SQ SEQUENCE 282 AA; 31169 MW; 9C8F6FF37D50BC09 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 282;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 245 FEFIMNDP 252

RESULT 70
Y079_RICCN STANDARD; PRT; 282 AA.
AC Q92JF8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Hypothetical protein RC0079.
GN RC0079.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT Mechanisms of evolution in Rickettsia conorii and R. prowazekii.;
RL Science 293:2093-2098(2001).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008575; AAL02617.1; -.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 282 AA; 31331 MW; 8C9CDD6E47065445 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 282;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 156 LEFIDKDG 163

RESULT 71
SPEE_CLOPE STANDARD; PRT; 283 AA.
AC Q8XMY8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Spermidine synthase (EC 2.5.1.16) (Putrescine aminopropyltransferase)
DE (SPDSY).
GN SPEE OR CPE0550.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
```

```
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RC MEDLINE=21664373; PubMed=11792842;
RX Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kohara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC -!- CATALYTIC ACTIVITY: S-adenosylmethioninamine + putrescine = 5'-
CC methylthioadenosine + spermidine.
CC -!- PATHWAY: Spermidine biosynthesis; fifth (last) step.
CC -!- SIMILARITY: Belongs to the spermidine/spermine synthase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP003187; BAB80256.1; -.
CC DR HAMAP; MF 00198; -.
CC DR InterPro; IPR000051; SAM bind.
CC DR InterPro; IPR001045; Spermine synthase.
CC DR Pfam; PF01564; Spermine synth, 1.
CC DR TIGRFAMs; TIGR00417; speE; 1.
CC DR PROSITE; PS01330; SPERMIDINE SYNTHASE; FALSE NEG.
CC KW Spermidine biosynthesis; Transference; Complete proteome.
CC FT DOMAIN 77 118 BINDING TO DECARBOXYLATED SAM
CC SQ SEQUENCE 283 AA; 32992 MW; F546286DA257EB22 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 283;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 57 DEFYHDM 64

RESULT 72
CELA_ACEXY STANDARD; PRT; 284 AA.
ID CELA_ACEXY
AC P27897;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) (UDP-glucose
DE pyrophosphorylase) (UDPGP) (Alpha-D-glucosyl-1-phosphate
DE uridylyltransferase) (uridine diphosphoglucose pyrophosphorylase).
GN CELA.
OS Acetobacter xylinus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Gluconacetobacter.
OX NCBI_TaxID=28448;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92041596; PubMed=1938907;
RA Brede G., Fjaervik E., Valla S.;
RT "Nucleotide sequence and expression analysis of the Acetobacter
RT xylinum uridine diphosphoglucose pyrophosphorylase gene.";
RL J. Bacteriol. 173:7042-7045(1991).
CC -!- CATALYTIC ACTIVITY: UTP + alpha-D-glucose 1-phosphate =
CC diphosphate + UDP-glucose.
CC -!- SIMILARITY: Belongs to the prokaryotic UDPGP family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M76548; AAA21888.1; -;  
 DR PIR; A41382; A41382.  
 DR InterPro; IPR005835; NTP transferase.  
 DR Pfam; PF00483; NTP transferase; 1.  
 DR Trnseq; Kinase; Nucleotidyltransferase.  
 KW SEQUENCE 284 AA; 30992 MW; 6C809AFFBBA1791B CRC64;

Query Match 100.0%; Score 25; DB 1; Length 284;  
 Best Local Similarity 50.0%; Pred. No. 3.2e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFTXXDX 8  
 :|||:|:  
 Db 120 REFIGNDP 127

## RESULT 73

RFBA\_NEIMA  
 ID\_RFBA\_NEIMA STANDARD; PRT; 288 AA.  
 AC P57040;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glucose-1-phosphate thymidyltransferase (EC 2.7.7.24) (dUDP-glucose  
 synthase)  
 GN (RFBAL OR NMA0198) AND (RFBAL OR NMA0205).  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699;  
 RN [1]

SEQUENCE FROM N.A.  
 RC STRAIN=22491 / Serogroup A / Serotype 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churche C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies K.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,  
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis 22491";  
 RL Nature 404:502-506(2000).  
 CC -!- CATALYTIC ACTIVITY: dTTP + alpha-D-glucose 1-phosphate =  
 CC diphosphate + dUDP-glucose.  
 CC -!- PATHWAY: RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN BIOSYNTHESIS  
 CC PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.  
 CC -!- SIMILARITY: Belongs to the glucose-1-phosphate  
 CC thymidyltransferase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AL162752; CAB83503.1; -;  
 DR EMBL; AL162752; CAB83519.1; -;  
 DR PIR; H82014; H82014.  
 DR InterPro; IPR005907; G1P thy trans 1.  
 DR InterPro; IPR005835; NTP transferase.  
 DR Pfam; PF00483; NTP transferase; 1.  
 DR TIGRFAMs; TIGR01207; rnaA; 1.  
 KW Lipopolysaccharide biosynthesis; Transferase; Kinase;  
 KW Nucleotidyltransferase; Complete proteome.

SQ SEQUENCE 288 AA; 32151 MW; C818727C546F5E00 CRC64;  
 Query Match 100.0%; Score 25; DB 1; Length 288;  
 Best Local Similarity 50.0%; Pred. No. 3.2e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFTXXDX 8  
 :|||:|:  
 Db 94 EEFIGNDN 101

## RESULT 74

RFBA\_NEIMB  
 ID\_RFBA\_NEIMB STANDARD; PRT; 288 AA.  
 AC P55255;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glucose-1-phosphate thymidyltransferase (EC 2.7.7.24) (dUDP-glucose  
 synthase)  
 GN (RFBAL OR NMB0062) AND (RFBAL OR NMB0080).  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN [1]

SEQUENCE FROM N.A.  
 RC STRAIN=B1940 / Serogroup B;  
 RX MEDLINE=94293762; PubMed=8022265;  
 RA Hammerschmidt S., Birkoiz C., Zahring U., Robertson B.D.,  
 RA van Putten J.P.M., Ebeling O., Froesch M.;  
 RT "Contribution of genes from the capsule gene complex (cps) to  
 lipopolysaccharide biosynthesis and serum resistance in Neisseria  
 meningitidis";  
 RL Mol. Microbiol. 11:885-896(1994).  
 RN [2]

SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / Serogroup B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Feden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Clecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,  
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 MC58";  
 RL Science 287:1809-1815(2000).  
 CC -!- CATALYTIC ACTIVITY: dTTP + alpha-D-glucose 1-phosphate =  
 CC diphosphate + dUDP-glucose.  
 CC -!- PATHWAY: RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN BIOSYNTHESIS  
 CC PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.  
 CC -!- SIMILARITY: Belongs to the glucose-1-phosphate  
 CC thymidyltransferase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L09189; AAC37050.1; ALT\_INIT.  
 DR EMBL; AE002365; AAF40530.1; -;  
 DR EMBL; AE002367; AAF40544.1; -;  
 DR PIR; B81240; B81240.  
 DR TIGR; NMB0062; -;  
 DR TIGR; NMB0080; -;  
 DR InterPro; IPR005907; G1P thy trans 1.  
 DR InterPro; IPR005835; NTP transferase.

```

DR Pfam: PF00483; NTP transferase; 1.
DR TIGRfams: TIGR01207; rmlA; 1.
KW Lipopolysaccharide biosynthesis; Transferase; Kinase;
KW Nucleotidyltransferase; Complete proteome.
FT CONFLICT 125 125 A -> R (IN REF. 1).
FT CONFLICT 198 198 T -> S (IN REF. 1).
SQ SEQUENCE 288 AA; 32161 MW; 8A65B50B531F2907 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 288;
Best Local Similarity 50.0%; Pred. No. 3.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 94 EEFIGNDN 101

RESULT 75
RMLA_STRMU STANDARD; PRT; 289 AA.
AC P95778;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucose-1-phosphate thymidyltransferase (EC 2.7.7.24) (dUDP-glucose
DE synthase) (dUDP-glucose pyrophosphorylase).
GN RMLA OR SMU.1461.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=XC;
RA Tsukioaka Y., Yamashita Y., Nakano Y., Oho T., Koga T.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carlson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- CATALYTIC ACTIVITY: dTTP + alpha-D-glucose 1-phosphate =
CC diphosphate + dUDP-glucose.
CC -!- PATHWAY: dUDP-L-rhamnose biosynthesis.
CC -!- SIMILARITY: Belongs to the glucose-1-phosphate
CC thymidyltransferase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: D78182; BAAL1247.1; -
DR EMBL: AE014978; RAN59119.1; -
DR InterPro: IPR005907; GLP_thy_trans 1.
DR InterPro: IPR005835; NTP_transferase.
DR Pfam: PF00483; NTP transferase; 1.
DR TIGRfams: TIGR01207; rmlA; 1.
KW Lipopolysaccharide biosynthesis; Transferase; Kinase;
KW Nucleotidyltransferase; Complete proteome.
SQ SEQUENCE 289 AA; 32295 MW; 5563650C07C00987 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 289;
Best Local Similarity 50.0%; Pred. No. 3.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 94 EEFIGNDN 101

RESULT 76
ARY2_MOUSE STANDARD; PRT; 290 AA.
AC P50295;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arylamine N-acetyltransferase 2 (EC 2.3.1.5) (Arylamide acetylase 2)
DE (N-acetyltransferase type 2) (NAT-2).
GN NAT2 OR AAC2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6, and A/J;
RX MEDLINE=91342604; PubMed=1875909;
RA Martell K.J., Vatsis K.P., Weber W.W.;
RT "Molecular genetic basis of rapid and slow acetylation in mice.";
RL Mol. Pharmacol. 40:218-227(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HeJ; TISSUE=Heart;
RA Hein D.W., Doll M.A.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=94379961; PubMed=7545952;
RA Kelly S.L., Sim E.;
RT "Arylamine N-acetyltransferase in Balb/c mice: identification of a
RT novel mouse isoenzyme by cloning and expression in vitro.";
RL Biochem. J. 302:347-353(1994).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=92382565; PubMed=1513324;
RA Martell K.J., Levy G.N., Weber W.W.;
RT "Cloned mouse N-acetyltransferases: enzymatic properties of expressed
RT Nat-1 and Nat-2 gene products.";
RL Mol. Pharmacol. 42:265-272(1992).
CC -!- FUNCTION: Participates in the detoxification of a plethora of
CC hydrazine and arylamine drugs. 2-aminofluorene and p-aminobenzoic
CC acid (PABA) are preferred substrates for NAT-2. Less activity with
CC anisidine and barely detectable with SM2.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + an arylamine = CoA + an N-
CC acetylarylamine.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- POLYMORPHISM: There are two forms of NAT2: a rapid/stable isoform
CC (Asn-99) and a slow/unstable isoform (Ile-99).
CC -!- SIMILARITY: Belongs to the arylamine N-acetyltransferase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: U35886; AAA78943.1; -
DR EMBL: U35887; AAA78944.1; -
DR EMBL: U37249; AAA80353.1; -
DR EMBL: U37250; AAA80354.1; -
DR PIR: B61267; B61267.
DR MGD; MGI:109201; Nat2.
DR InterPro: IPR004081; AANacetyltransf.

```

```

DR InterPro; IPR001447; Acetyltransferase.
DR Pfam; PF00797; Acetyltransferase2; 1.
DR PRINTS; PR01543; ANATRSFRASE.
DR ProDom; PD002783; Acetyltransferase2; 1.
DR Transferase; Acyltransferase; Multigene family; Polymorphism.
FT ACT SITE 68 68 BY SIMILARITY.
FT VARIANT 99 99 N -> I (in allele NAT2*9; slow/unstable isoform).
SQ SEQUENCE 290 AA; 33701 MW; 704E000DB48CE557 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 290;
Best Local Similarity 50.0%; Pred. No. 3.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 173 QEFINDSL 180

RESULT 77
ENG STAAAC STANDARD; PRT; 291 AA.
AC Q9KX08;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Probable GTPase engc (EC 3.6.1.-)
GN ENGC
OS Staphylococcus aureus (strain COL)
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=93062;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=20031141; PubMed=10566865;
RX De Lencastre H., Wu S.-W., Pinho M.G., Ludovice A.M., Felipe S., Gardete S., Sobral R., Gill S., Chung M., Tomasz A.; "Antibiotic resistance as a stress response: complete sequencing of a large number of chromosomal loci in Staphylococcus aureus strain COL that impact on the expression of resistance to methicillin."; Microb. Drug Resist. 5:163-175(1999).
RL Microb. Drug Resist. 5:163-175(1999).
CC -!- FUNCTION: Unusual circularly permuted GTPase that catalyzes rapid hydrolysis of GTP with a slow catalytic turnover (By similarity).
CC -!- SUBUNIT: Monomer (Probable).
CC -!- SIMILARITY: Contains 1 engc GTPase domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
DR EMBL; Y13639; CAA73981.1; -.
DR HAMAP; MF_01820; -.
DR InterPro; IPR004861; DUF258.
DR Pfam; PF03193; DUF258; 1.
DR TIGRFAMs; TIGR00157; TIGR00157; 1.
DR PROSITE; PS50936; ENGC_GTPASE; 1.
DR Hydrolase; GTP-binding.
FT DOMAIN 72 219 ENGC GTPASE.
FT NP_BIND 112 115 GTP (PROBABLE).
FT NP_BIND 164 171 GTP (PROBABLE).
FT NP_BIND 214 218 GTP (PROBABLE).
FT SITE 245 258 KNUCKLE-LIKE CYSTEINE CLUSTER.
SQ SEQUENCE 291 AA; 33890 MW; 3D7FE2B5989577D6 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 291;
Best Local Similarity 50.0%; Pred. No. 3.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8

```

```

Db 139 TEFIGNDD 146

RESULT 78
ENG STAAAC STANDARD; PRT; 291 AA.
AC Q99UP7;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Probable GTPase engc (EC 3.6.1.-)
GN ENGC OR SAV1221 OR SA1064 OR MW1104;
OS Staphylococcus aureus (strain MU50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 196620;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=MU50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I., Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-O., Ito T., Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hoshoyama A., Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C., Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.; "Whole genome sequencing of methicillin-resistant Staphylococcus aureus."; Lancet 357:1225-1240(2001).
RL Lancet 357:1225-1240(2001).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=MW2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A., Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L., Yamamoto K., Hiramatsu K.; "Genome and virulence determinants of high virulence community-acquired MRSA."; Lancet 359:1819-1827(2002).
RL Lancet 359:1819-1827(2002).
CC -!- FUNCTION: Unusual circularly permuted GTPase that catalyzes rapid hydrolysis of GTP with a slow catalytic turnover (By similarity).
CC -!- SUBUNIT: Monomer (Probable).
CC -!- SIMILARITY: Contains 1 engc GTPase domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
DR EMBL; AP003361; BAB57383.1; -.
DR EMBL; AP003133; BAB42316.1; -.
DR EMBL; AP004826; BAB94969.1; -.
DR PIR; H89894; H89894.
DR HAMAP; MF_01820; -.
DR InterPro; IPR004881; DUF258.
DR Pfam; PF03193; DUF258; 1.
DR TIGRFAMs; TIGR00157; TIGR00157; 1.
DR PROSITE; PS50936; ENGC_GTPASE; 1.
DR Hydrolase; GTP-binding; Complete proteome.
FT DOMAIN 72 219 ENGC GTPASE.
FT NP_BIND 112 115 GTP (PROBABLE).
FT NP_BIND 164 171 GTP (PROBABLE).
FT NP_BIND 214 218 GTP (PROBABLE).
FT SITE 245 258 KNUCKLE-LIKE CYSTEINE CLUSTER.
SQ SEQUENCE 291 AA; 33875 MW; 27E10EFD86FB3263 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 291;
Best Local Similarity 50.0%; Pred. No. 3.2e+02;

```



DT 01-OCT-1994 (Rel. 30, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glucose-1-phosphate thymidyltransferase (EC 2.7.7.24) (dUDP-glucose  
 DE synthase) (dUDP-glucose pyrophosphorylase).  
 GN RFA OR SF2102 OR S2225.  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Shigella.  
 OX NCBI\_TaxID=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PR577 / Serotype 2a;  
 RX MEDLINE=94224146; PubMed=8170390;  
 RA Macpherson D.F., Manning P.A., Morona R.;  
 RT "Characterization of the dUDP-rhamnose biosynthetic genes encoded in  
 RT the rfb locus of Shigella flexneri.";  
 RL Mol. Microbiol. 11:281-292 (1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=301 / Serotype 2a;  
 RX MEDLINE=22272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
 RT through comparison with genomes of Escherichia coli K12 and O157";  
 RL Nucleic Acids Res. 30:4432-4441 (2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;  
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RT "Complete genome sequence and comparative genomics of Shigella  
 RT flexneri serotype 2a strain 2457T";  
 RL Infect. Immun. 71:2775-2786 (2003).  
 CC -!- CATALYTIC ACTIVITY: dUDP + alpha-D-glucose 1-phosphate =  
 CC diphosphate + dUDP-glucose.  
 CC -!- PATHWAY: RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN BIOSYNTHESIS  
 CC PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.  
 CC -!- SIMILARITY: Belongs to the glucose-1-phosphate  
 CC thymidyltransferase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X71970; CAA50769.1; -;  
 CC EMBL; L14842; AAA33681.1; -;  
 CC EMBL; AB015225; AAN43641.1; -;  
 CC EMBL; AB016985; AAP17470.1; -;  
 CC PIR; S41536; S41536;  
 CC InterPro; IPR005907; GLP\_thy\_trans.1.  
 CC InterPro; IPR005835; NTP\_transferase.  
 CC Pfam; PF00483; NTP\_transferase; 1.

DR TIGRfams: TIGR01207; rmlA; 1.  
 KW Lipopolysaccharide biosynthesis; Transferase; Kinase;  
 KW Nucleotidyltransferase; Complete proteome.  
 FT CONFLICT 157 157 A -> P (IN REF. 1).  
 FT CONFLICT 166 166 E -> Q (IN REF. 1).  
 FT CONFLICT 215 215 MISSING (IN REF. 1).  
 FT CONFLICT 224 224 W -> C (IN REF. 1).  
 FT CONFLICT 228 233 GTHQSL -> DTSKP (IN REF. 1).  
 FT CONFLICT 243 245 IEE -> NED (IN REF. 1).  
 SQ SEQUENCE 292 AA; 32487 MW; E54F6199E0361AF2 CRC64;  
 Query Match 100.0%; Score 25; DB 1; Length 292;  
 Best Local Similarity 50.0%; Pred. No. 3.3e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 XEFIXDX 8  
 Db :|||:|:  
 97 EEFIGDD 104  
 RESULT 82  
 ID RBAL ECOLI STANDARD; PRT; 293 AA.  
 AC P37744; P78081;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glucose-1-phosphate thymidyltransferase (EC 2.7.7.24) (dUDP-glucose  
 DE synthase) (dUDP-glucose pyrophosphorylase).  
 GN RFA OR RMLA OR B2039.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / WGI;  
 RX MEDLINE=94292435; PubMed=7517391;  
 RA Stevenson G., Neal B., Liu D., Hobbs M., Packer N.H., Batley M.,  
 RA Redmond J.W., Lindquist L., Reeves P.R.;  
 RT "Structure of the O antigen of Escherichia coli K-12 and the sequence  
 RT of its rfb gene cluster";  
 RL J. Bacteriol. 176:4144-4156 (1994).  
 RN [2]  
 RP REVISION TO 288.  
 RC STRAIN=K12 / WGI;  
 RA Stevenson G.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474 (1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251358; PubMed=9097040;  
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,  
 RA Makino K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,  
 RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,  
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;  
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 40.1-50.0 min region on the linkage map";  
 RL DNA Res. 3:379-392 (1996).  
 RN [5]  
 RP SEQUENCE OF 247-293 FROM N.A.



```

RC STRAIN=K12 / W3110;
RX MEDLINE=94292434; PubMed=7517390;
RA Yao Z., Valvano M.A.;
RT "Genetic analysis of the O-specific lipopolysaccharide biosynthesis
RT region (rfb) of Escherichia coli K-12 W3110: identification of genes
RT that confer group 6 specificity to Shigella flexneri serotypes Y and
RT 4a.";
RL J. Bacteriol. 176:4133-4143(1994).
CC -1- CATALYTIC ACTIVITY: dGTP + alpha-D-glucose 1-phosphate =
CC diphosphate + dNDP-glucose.
CC -1- PATHWAY: dTDP-L-RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
CC -1- SIMILARITY: Belongs to the glucose-1-phosphate
CC thymidyltransferase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U09876; AAB88400.1; -.
CC EMBL; AB000294; AAC75100.1; -.
CC EMBL; D90841; BAA15881.1; -.
CC EMBL; D90842; BAA15893.1; -.
CC EMBL; U03041; AAC31629.1; -.
CC PIR; P64969; P64969.
CC PDB; 1H5T; 08-MAY-02.
CC SWISS-2DPAGE; P37744; COLI.
CC EcoGene; Egl1978; rfbA.
DR InterPro; IPR005907; GIP_thy_trans_1.
DR InterPro; IPR005835; NTP_transferase.
DR Pfam; PF00483; NTP_transferase; 1.
DR TIGRFAMs; TIGR01207; rmlA; 1.
KW Lipopolysaccharide biosynthesis; Transferase; Kinase;
KW Nucleotidyltransferase; Complete proteome; 3D-structure.
FT CONFLICT 247 247 Q -> P (IN REF. 5).
SQ SEQUENCE 293 AA; 32693 MW; BA895362D1C5CA55 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 293;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIGDDX 8
Db |||:::
97 BEFIGDD 104

RESULT 83
FTR METKA
ID FTR METKA STANDARD; PRT; 296 AA.
AC Q49610;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Formylmethanofuran--tetrahydromethanopterin formyltransferase
DE (EC 2.3.1.101) (HAMPT formyltransferase).
GN FTR OR MK0116.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=95324581; PubMed=7601152;
RA Shima S., Weiss D.S., Thauer R.K.;
RT "Formylmethanofuran:tetrahydromethanopterin formyltransferase (Ftr)
RT from the hyperthermophilic Methanopyrus kandleri. Cloning, sequencing
RT and functional expression of the ftr gene and one-step purification
RT of the enzyme overproduced in Escherichia coli.";
KW

```

```

RL Eur. J. Biochem. 230:906-913(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatsov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozlyavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
RN [3]
RP SEQUENCE OF 1-49, AND CHARACTERIZATION.
RX MEDLINE=931130924; PubMed=1483480;
RA Breitung J., Borner G., Scholz S., Linder D., Stetter K.O.,
RA Thauer R.K.;
RT "Salt dependence, kinetic properties and catalytic mechanism of N-
RT formylmethanofuran:tetrahydromethanopterin formyltransferase from the
RT extreme thermophile Methanopyrus kandleri.";
RL Eur. J. Biochem. 210:971-981(1992).
RN [4]
RP SUBUNIT, AND MUTAGENESIS OF ARG-261.
RX MEDLINE=20507566; PubMed=11054114;
RA Shima S., Thauer R.K., Ermier U., Durchschlag H., Tziatzios C.,
RA Schubert D.;
RT "A mutation affecting the association equilibrium of
RT formyltransferase from the hyperthermophilic Methanopyrus kandleri
RT and its influence on the enzyme's activity and thermostability.";
RL Eur. J. Biochem. 267:6619-6623(2000).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE=97341227; PubMed=9195883;
RA Ermier U., Mörchel M., Thauer R., Shima S.;
RT "Formylmethanofuran: tetrahydromethanopterin formyltransferase from
RT Methanopyrus kandleri - new insights into salt-dependence and
RT thermostability.";
RL Structure 5:635-646(1997).
CC -1- FUNCTION: Catalyzes the reversible transfer of a formyl group from
CC formylmethanofuran (formyl-MFR) to tetrahydromethanopterin
CC (H(4)MPT) so as to produce 5-formyl-methanofuran + 5,6,7,8-
CC CATALYTIC ACTIVITY: N-formylmethanofuran + 5,6,7,8-
CC tetrahydromethanopterin = methanofuran + 5-formyl-5,6,7,8-
CC tetrahydromethanopterin.
CC -1- ENZYME REGULATION: Requires high salt concentrations for
CC thermostability.
CC -1- PATHWAY: Methanogenesis from carbon dioxide; second step.
CC -1- SUBUNIT: Homotetramer composed of two dimers. Dimerization is
CC sufficient for enzyme activity, but tetramerization is required
CC for high thermostability.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the FTR family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X85115; CAA59435.1; -.
CC EMBL; AB010311; AAM01333.1; -.
CC PIR; S65950; S57647.
CC PDB; 1FTR; 25-NOV-98.
CC HAMAP; MF_00579; -.
CC InterPro; IPR002770; FTR.
CC Pfam; PF01913; FTR; 1.
CC Pfam; PF02741; FTR; 1.
CC PIRSF; PIRSF006414; FTR; 1.
CC ProDom; PD007702; FTR; 1.
CC Methanogenesis; One-carbon metabolism; Transferase; Acyltransferase;
KW

```



```

KW 3D-structure; Complete proteome.
FT MUTAGEN 261 261 R->E: WEAKENS DIMER-DIMER ASSOCIATION.
FT CONFLICT 30 30 H -> D (IN REF. 3).
FT CONFLICT 32 32 W -> K (IN REF. 3).
FT CONFLICT 39 39 E -> K (IN REF. 3).
FT STRAND 2 3
FT TURN 4 5
FT STRAND 6 8
FT STRAND 12 26
FT HELIX 30 41
FT TURN 47 49
FT STRAND 53 61
FT HELIX 63 65
FT TURN 67 68
FT STRAND 72 79
FT HELIX 82 96
FT TURN 97 99
FT TURN 101 102
FT STRAND 104 107
FT HELIX 111 113
FT STRAND 116 118
FT HELIX 120 125
FT HELIX 126 128
FT TURN 129 130
FT STRAND 133 137
FT TURN 138 139
FT STRAND 140 147
FT TURN 148 149
FT STRAND 150 155
FT STRAND 158 173
FT HELIX 176 190
FT TURN 191 192
FT TURN 194 195
FT STRAND 196 197
FT HELIX 201 203
FT STRAND 205 205
FT STRAND 209 211
FT TURN 217 218
FT STRAND 221 223
FT HELIX 225 227
FT TURN 229 230
FT TURN 232 233
FT TURN 238 239
FT STRAND 242 250
FT HELIX 253 267
FT TURN 268 269
FT TURN 271 272
FT STRAND 273 277
FT STRAND 282 283
FT STRAND 288 291
FT HELIX 292 296
SQ SEQUENCE 296 AA; DDE02D3E7D98FC86 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 296;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 150 GBFVDS 157

RESULT 84
HSLO CLOAB STANDARD; PRT; 297 AA.
AC Q97GJ6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 33 kDa chaperonin (Heat shock protein 33 homolog) (HSP33).
GN HSLO OR CAC2370.
OS Clostridium acetobutylicum.

KW 3D-structure; Complete proteome.
FT MUTAGEN 261 261 R->E: WEAKENS DIMER-DIMER ASSOCIATION.
FT CONFLICT 30 30 H -> D (IN REF. 3).
FT CONFLICT 32 32 W -> K (IN REF. 3).
FT CONFLICT 39 39 E -> K (IN REF. 3).
FT STRAND 2 3
FT TURN 4 5
FT STRAND 6 8
FT STRAND 12 26
FT HELIX 30 41
FT TURN 47 49
FT STRAND 53 61
FT HELIX 63 65
FT TURN 67 68
FT STRAND 72 79
FT HELIX 82 96
FT TURN 97 99
FT TURN 101 102
FT STRAND 104 107
FT HELIX 111 113
FT STRAND 116 118
FT HELIX 120 125
FT HELIX 126 128
FT TURN 129 130
FT STRAND 133 137
FT TURN 138 139
FT STRAND 140 147
FT TURN 148 149
FT STRAND 150 155
FT STRAND 158 173
FT HELIX 176 190
FT TURN 191 192
FT TURN 194 195
FT STRAND 196 197
FT HELIX 201 203
FT STRAND 205 205
FT STRAND 209 211
FT TURN 217 218
FT STRAND 221 223
FT HELIX 225 227
FT TURN 229 230
FT TURN 232 233
FT TURN 238 239
FT STRAND 242 250
FT HELIX 253 267
FT TURN 268 269
FT TURN 271 272
FT STRAND 273 277
FT STRAND 282 283
FT STRAND 288 291
FT HELIX 292 296
SQ SEQUENCE 296 AA; DDE02D3E7D98FC86 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 296;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 150 GBFVDS 157

RESULT 85
VPO_BPH1 STANDARD; PRT; 298 AA.
AC P51719;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Probable capsid scaffolding protein (ORF17).
OS Bacteriophage HPI.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.
OX NCBI_TaxID=10690;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HPIc1;
RX MEDLINE=96279738; PubMed=8710508;
RA Esposito D., Fitzmaurice W.P., Benjamin R.C., Goodman S.D.,
RA Waldman A.S., Socca J.J.;
RT "The complete nucleotide sequence of bacteriophage HPI DNA.";
RL Nucleic Acids Res 24:2360-2368(1996).
CC -!- SIMILARITY: STRONG, TO PHAGE P2 PROTEIN O.

Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
NCBI_TaxID=1488;
[1]
SEQUENCE FROM N.A.
STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
MEDLINE=21359325; PubMed=1146286;
Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
Bennett G.N., Koonin E.V., Smith D.R.;
"Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
J. Bacteriol. 183:4823-4838(2001).
-!- FUNCTION: Redox regulated molecular chaperone. Protects both
thermally unfolding and oxidatively damaged proteins from
irreversible aggregation. Plays an important role in the bacterial
defense system toward oxidative stress (By similarity).
-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
-!- PFM: Under oxidizing conditions two disulfide bonds are formed
involving the reactive cysteines. Under reducing conditions zinc
is bound to the reactive cysteines and the protein is inactive (By
similarity).
-!- SIMILARITY: Belongs to the HSP33 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AE007737; AAK80326.1; -
PIR; C97192; C97192.
HAMAP; ME 00117; -; 1.
InterPro; IPR000397; Hsp33.
Pfam; PF01430; HSP33; 1.
ProDom; PD248154; Hsp33; 1.
Chaperone; Redox-active center; Zinc; Complete proteome.
FT DISULFID 239 241 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 272 275 REDOX-ACTIVE (BY SIMILARITY).
SQ SEQUENCE 297 AA; 32033 MW; 4E3B6452107316D CRC64;

Query Match 100.0%; Score 25; DB 1; Length 297;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 217 LEFIFEDM 224

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U24159; AAB09202.1; -.
DR PIR; S69523; S69523.
KW Capsid assembly.
SQ SEQUENCE 298 AA; 33702 MW; 6D2841468D700C1F CRC64;
Query Match 100.0%; Score 25; DB 1; Length 298;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db 142 GEFIKVDF 149
RESULT 86
YL44_AQUAE STANDARD; PRT; 298 AA.
AC O67900;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_2144.
GN AQ_2144.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Auja M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000775; AAC07871.1; -.
DR PIR; H70483; H70483.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 298 AA; 34256 MW; EBD06E78B3D338EF CRC64;
Query Match 100.0%; Score 25; DB 1; Length 298;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db 227 BEFILEDF 234
RESULT 87
YL4IF_RHISN STANDARD; PRT; 305 AA.
ID Y4IF_RHISN
AC P55489;
DT 01-NOV-1997 (Rel. 35, Created)

```

```

DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Hypothetical 34.1 kDa protein Y4IF.
GN Rhizobium sp. (strain NGR234).
OS Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 397:394-401(1997).
CC -!- SIMILARITY: None obvious.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000078; AAB91701.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 305 AA; 34141 MW; 331117881829F351 CRC64;
Query Match 100.0%; Score 25; DB 1; Length 305;
Best Local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 XEFIXDX 8
Db 113 IEFIRGDI 120
RESULT 88
COAA_LACLA STANDARD; PRT; 306 AA.
ID COAA_LACLA
AC Q9CFM3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pantothenate kinase (EC 2.7.1.33) (Pantothenic acid kinase).
GN COAA Or Lbl444.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403";
RL Genome Res. 11:731-753(2001).
CC -!- CATALYTIC ACTIVITY: ATP + pantothenate = ADP + D-4'-
CC phosphopantothenate.
CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the prokaryotic pantothenate kinase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

CC EMBL; AE006374; AAK05542.1; -;  
 DR PIR; D86805; D86805.  
 DR HAMAP; MF 00215; -; 1.  
 DR InterPro; IPR004566; Pank\_bact.  
 DR InterPro; IPR006083; PRK\_URK.  
 DR Pfam; PF00485; PRK; 1.  
 DR PIRSF; PIRSF00545; Pantothenate\_kin; 1.  
 KW Transferase; Kinase; ATP-binding; Coenzyme A biosynthesis;  
 KW Complete proteome.  
 FT NP BIND 90 97 ATP (POTENTIAL).  
 SQ SEQUENCE 306 AA; 35992 MW; 87241BDFC3FC9D8 CRC64;  
 Query Match 100.0%; Score 25; DB 1; Length 306;  
 Best Local Similarity 50.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEFIXDX 8  
 Db :|||:|:  
 2 NEFINFDE 9  
 RESULT 89  
 COA STRA3  
 ID COA STRA3 STANDARD; PRT; 306 AA.  
 AC Q8E5P2; Q8D2Z1;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Pantothenate kinase (EC 2.7.1.33) (Pantothenic acid kinase).  
 GN COA OR GBS0939 OR SAG0951.  
 OS Streptococcus agalactiae (serotype III), and  
 OS Streptococcus agalactiae (serotype V).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216495, 216466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaser P., Rusniok C., Buchrieser C., Chevallier F., Frangeul L.,  
 RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease."  
 RL Mol. Microbiol. 45:1499-1513(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2603 V/R / Serotype V;  
 RX MEDLINE=2222988; PubMed=12200547;  
 RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,  
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,  
 RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,  
 RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,  
 RA Radu D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,  
 RA Carthy H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,  
 RA Iacobini E.T., Brettoni C., Galli G., Mairani M., Vegni F., Maione D.,  
 RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative genomic analysis of an  
 RT emerging human pathogen, serotype V Streptococcus agalactiae."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).  
 CC -!- CATALYTIC ACTIVITY: ATP + pantothenate = ADP + D-4'-  
 CC phosphopantothenate.  
 CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; first step.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -!- SIMILARITY: Belongs to the prokaryotic pantothenate kinase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sb-sib.ch).  
 CC -----  
 DR EMBL; AL766848; CAD46598.1; -;  
 DR EMBL; AE014235; AAM99835.1; -;  
 DR SgaList; gbs0939; -;  
 DR TIGR; SAG0951; -;  
 DR HAMAP; MF 00215; -; 1.  
 DR InterPro; IPR004566; Pank\_bact.  
 DR InterPro; IPR006083; PRK\_URK.  
 DR Pfam; PF00485; PRK; 1.  
 DR PIRSF; PIRSF00545; Pantothenate\_kin; 1.  
 KW Transferase; Kinase; ATP-binding; Coenzyme A biosynthesis;  
 KW Complete proteome.  
 FT NP BIND 91 98 ATP (POTENTIAL).  
 SQ SEQUENCE 306 AA; 36093 MW; 01FF015134D76D34 CRC64;  
 Query Match 100.0%; Score 25; DB 1; Length 306;  
 Best Local Similarity 50.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 XEFIXDX 8  
 Db :|||:|:  
 3 NEFINFDR 10  
 RESULT 90  
 OPFF\_BACSU  
 ID OPFF\_BACSU STANDARD; PRT; 308 AA.  
 AC P24137; O31599; P23366;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Oligopeptide transport ATP-binding protein oppF.  
 GN OPFF OR SPO0KE OR BSU11470.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=91194547; PubMed=1901616;  
 RA Rudner D.Z., Ledoux J.R., Ireton K., Grossman A.D.;  
 RT "The spo0K locus of Bacillus subtilis is homologous to the  
 RT oligopeptide permease locus and is required for sporulation and  
 RT competence."  
 RL J. Bacteriol. 173:1388-1398(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Chou S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Enian K.P., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Gham S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

```

RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosoato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler H., Weitzenger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RA "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256 (1997).
CC -1- FUNCTION: Component of the oligopeptide permease, a binding
CC protein-dependent transport system. Necessary for genetic
CC competence but not sporulation. Probably responsible for energy
CC coupling to the transport system.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- SIMILARITY: Belongs to the ABC transporter family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X56347; CAA39791.1; -.
CC EMBL; M57689; AAA62692.1; -.
CC EMBL; M57689; AAA62693.1; ALT INIT.
CC EMBL; Z99110; CAB13004.1; ALT_INIT.
CC PIR; E38447; E38447.
CC Subtilist; BGL0775; opf.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003439; ABC_transporter.
CC Pfam; PF00005; ABC_tran; 1.
CC ProDom; PD000006; ABC_transporter; 1.
CC SMART; SM00382; AAA; 1.
CC PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
CC PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
CC Competence; Peptide transport; Transport; Membrane; ATP-binding;
CC Sporulation; Complete proteome.
CC NP BIND 45 52 ATP (BY SIMILARITY).
CC FT CONFLICT 269 272 VRQK -> CSE (IN REF. 1).
CC SEQUENCE 308 AA; 35104 MW; F17DE0016AA3E4ED CRC64;

Query Match 100.0%; Score 25; DB 1; Length 308;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 172 PEFIAD 179
:||||:
:||||:

RESULT 91
MDH_CHLAU STANDARD; PRT; 309 AA.
ID -MDH_CHLAU
AC P80040;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
DE MDH.
GN MDH.
OS Chloroflexus aurantiacus.
OC Bacteria; Chloroflexi; Chloroflexales; Chloroflexaceae; Chloroflexus.
OX NCBI_TaxID=1108;
RN [1]

RA Medina N., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosoato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler H., Weitzenger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RA "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256 (1997).
CC -1- FUNCTION: Component of the oligopeptide permease, a binding
CC protein-dependent transport system. Necessary for genetic
CC competence but not sporulation. Probably responsible for energy
CC coupling to the transport system.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- SIMILARITY: Belongs to the ABC transporter family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X56347; CAA39791.1; -.
CC EMBL; M57689; AAA62692.1; -.
CC EMBL; M57689; AAA62693.1; ALT INIT.
CC EMBL; Z99110; CAB13004.1; ALT_INIT.
CC PIR; E38447; E38447.
CC Subtilist; BGL0775; opf.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003439; ABC_transporter.
CC Pfam; PF00005; ABC_tran; 1.
CC ProDom; PD000006; ABC_transporter; 1.
CC SMART; SM00382; AAA; 1.
CC PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
CC PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
CC Competence; Peptide transport; Transport; Membrane; ATP-binding;
CC Sporulation; Complete proteome.
CC NP BIND 45 52 ATP (BY SIMILARITY).
CC FT CONFLICT 269 272 VRQK -> CSE (IN REF. 1).
CC SEQUENCE 308 AA; 35104 MW; F17DE0016AA3E4ED CRC64;

Query Match 100.0%; Score 25; DB 1; Length 308;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 172 PEFIAD 179
:||||:
:||||:

RESULT 92
ANXC_HYDAT STANDARD; PRT; 316 AA.
ID -ANXC_HYDAT
AC P26256;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Annexin B12 (Annexin XII).
DE ANXB12.
GN ANXB12.
OS Hydra attenuata (Hydra) (Hydra vulgaris).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_TaxID=6087;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92250599; PubMed=1339458;
RA Schlaepfer D.D., Fisher D.A., Brandt M.E., Bode H.R., Jones J.M.,
RA Haigler H.T.;
RA "Identification of a novel annexin in Hydra vulgaris.
RT
```

RT Characterization, cDNA cloning, and protein kinase C phosphorylation  
 of annexin XII.";  
 RL J. Biol. Chem. 267:9529-9539(1992).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE=96085034; PubMed=7477411;  
 RA Luecke H., Chang B.T., Maillard W.S., Schlaepfer D.D., Haigler H.T.;  
 RT "Crystal structure of the annexin XII hexamer and implications for  
 RL bilayer insertion.";  
 RL Nature 378:512-515(1995).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.93 ANGSTROMS) OF MUTANT LYS-105.  
 RX MEDLINE=20170873; PubMed=10704197;  
 RA Cartallier J.P., Haigler H.T., Luecke H.;  
 RT "Annexin XII E105K crystal structure: identification of a pH-dependent  
 RL switch for mutant hexamerization.";  
 RL Biochemistry 39:2475-2483(2000).  
 CC -1- SUBUNIT: Homohexamer.  
 CC -1- DOMAIN: A pair of annexin repeats may form one binding site for  
 CC calcium and phospholipid.  
 CC -1- PTM: Phosphorylated in vitro on serine(s) and threonine(s) by PKC.  
 CC -1- SIMILARITY: Belongs to the annexin family.  
 CC -1- SIMILARITY: Contains 4 annexin repeats.  
 CC -1- DATABASE: NMR-Annexin 12 home page from Luecke's group;  
 CC WWW="http://anx12.bio.ucl.edu/~hudel/anx12/".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M83736; AAA23206.1; -;  
 DR PIR; A42660; LUUF12.  
 DR PDB; LAEI; 20-AUG-97.  
 DR PDB; LDM5; 20-MAR-00.  
 DR InterPro; IPR001464; Annexin.  
 DR Pfam; PF00191; annexin; 4.  
 DR PRINTS; PR00196; ANNEXIN.  
 DR ProDom; PD000143; Annexin; 4.  
 DR SMART; SM00335; ANX; 4.  
 DR PROSITE; PS00223; ANNEXIN; 3.  
 KW Annexin; Calcium/phospholipid-binding; Repeat; Phosphorylation;  
 3D-structure.  
 FT REPEAT 22 82 ANNEXIN 1.  
 FT REPEAT 94 154 ANNEXIN 2.  
 FT REPEAT 178 238 ANNEXIN 3.  
 FT REPEAT 253 313 ANNEXIN 4.  
 FT MOD\_RES 6 6 PHOSPHORYLATION (BY PKC) (POTENTIAL).  
 FT TURN 11 12  
 FT HELIX 15 26  
 FT HELIX 33 42  
 FT HELIX 45 59  
 FT HELIX 63 70  
 FT HELIX 73 83  
 FT HELIX 86 99  
 FT HELIX 105 112  
 FT HELIX 117 131  
 FT HELIX 135 142  
 FT HELIX 145 155  
 FT TURN 156 157  
 FT TURN 161 162  
 FT HELIX 167 180  
 FT TURN 181 183  
 FT HELIX 189 198  
 FT HELIX 201 214  
 FT HELIX 219 226  
 FT HELIX 229 257  
 FT TURN 259 260  
 FT HELIX 264 274  
 FT TURN 275

FT HELIX 279 290  
 FT HELIX 294 301  
 FT HELIX 304 314  
 SQ SEQUENCE 316 AA; 35108 MW; 10599869CBA853EC CRC64;  
 Query Match 100.0%; Score 25; DB 1; Length 316;  
 Best Local Similarity 50.0%; Pred. No. 3.6e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 XEFIXDX 8  
 Db 295 YEFITDDC 302  
 :|||:|:  
 RESULT 93  
 RPOA\_AQUAE STANDARD; PRT; 317 AA.  
 ID RPOA\_AQUAE  
 AC O66483;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (RNAP alpha  
 DE subunit) (Transcriptase alpha chain) (RNA polymerase alpha subunit).  
 GN RPOA OR AQ 070.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 OX NCBI\_TaxID=63363;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aulay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus.";  
 RL Nature 392:353-358(1998).  
 CC -1- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
 CC of DNA into RNA using the four ribonucleoside triphosphates as  
 CC substrates.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: Homodimer. The RNAP catalytic core consists of 2 alpha, 1  
 CC beta, 1 beta' and 1 omega subunit (By similarity).  
 CC -1- DOMAIN: The N-terminal domain is essential for RNAP assembly and  
 CC basal transcription, whereas the C-terminal domain is involved in  
 CC interaction with transcriptional regulators and with upstream  
 CC promoter elements (By similarity).  
 CC -1- SIMILARITY: Belongs to the RNA polymerase alpha chain family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000672; AAC06440.1; -;  
 DR PIR; H70306; H70306.  
 DR HSSP; Q9KWJ8; 1HQW.  
 DR HAMAP; MF\_00059; -; 1.  
 DR InterPro; IPR009025; RBP11-like\_RNAPo.  
 DR InterPro; IPR001700; RNA\_pola\_bac\_org.  
 DR Pfam; PF01000; RNA\_pola\_bac\_1.  
 DR Pfam; PF03118; RNA\_pola\_bac\_1.  
 DR ProDom; PD001179; RNA\_pola\_bac.org; 1.  
 DR SMART; SM00662; RPOLD; 1.  
 KW Transferase; Transcription; DNA-directed RNA polymerase;  
 Complete proteome.  
 FT DOMAIN 1 229 ALPHA N-TERMINAL DOMAIN (ALPHA-NTD)  
 FT (BY SIMILARITY).  
 FT DOMAIN 245 317 ALPHA C-TERMINAL DOMAIN (ALPHA-CTD)

FT SQ SEQUENCE 317 AA; 35768 MW; BC6EB9015163335 CRC64;  
(BY SIMILARITY).

Query Match 100.0%; Score 25; DB 1; Length 317;  
Best Local Similarity 50.0%; Pred. No. 3.6e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 3 NEFIYDK 10

## RESULT 94

CYF\_PSIU STANDARD; PRT; 321 AA.  
AC Q8W107;  
DT 28-FEB-2003 (Rel. 41, Last created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Apocytochrome f precursor.  
GN PETA.  
OS Psilotum nudum (Whisk fern).  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Moniliformopses; Psilotophyta; Psilotales; Psilotaceae; Psilotum.  
OX NCBI\_TaxID=3240;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Kinyok; Yamada K., Sugura M.;  
RA "Complete nucleotide sequence of the chloroplast genome from a fern,  
RT Psilotum nudum.";  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Translocates protons across the thylakoid membrane and  
transfers electrons from photosystem II to photosystem I. It  
receives electrons from the Rieske iron-sulfur protein and passes  
them to plastocyanin.  
CC -!- SUBUNIT: Interacts with plastocyanin and Rieske iron-sulfur  
protein (By similarity).  
CC -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane. Membrane-  
anchored (By similarity).  
CC -!- SIMILARITY: Belongs to the cytochrome c family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AP004638; BAB84229.1; -.  
DR HAMAP; MF 00610; -; 1.  
DR InterPro; IPR002325; Apocyt F.  
DR InterPro; IPR000345; CytC heme BS.  
DR Pfam; PF01333; Apocytochr F C; 1.  
DR PRINTS; PR00610; CYTOCHROME F.  
DR PROSITE; PS00190; CYTOCHROME C; 1.  
KW Electron transport; Heme; Chloroplast; Thylakoid; Photosynthesis;  
KW Photosystem I; Photosystem II; Transmembrane; Signal.  
FT SIGNAL 1 35  
FT CHAIN 36 321 APOCYTOCHROME F.  
FT TRANSMEM 287 306  
FT METAL 36 36  
FT IRON (HEME AXIAL LIGAND) (VIA AMINO  
NITROGEN) (BY SIMILARITY).  
FT BINDING 56 56 HEME (COVALENT) (BY SIMILARITY).  
FT BINDING 59 59 HEME (COVALENT) (BY SIMILARITY).  
FT METAL 60 60 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
SQ SEQUENCE 321 AA; 35893 MW; 2199A77B40B15428 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 321;  
Best Local Similarity 50.0%; Pred. No. 3.6e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 254 GEFIKIDQ 261

## RESULT 95

FCE2\_HUMAN STANDARD; PRT; 321 AA.  
AC P06734;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Low affinity immunoglobulin epsilon FC receptor (Lymphocyte IgE  
receptor) (Fc-epsilon-RII) (CD23) (BLAST-2) (Immunoglobulin E-binding  
factor).  
DE FCB2 OR IGEBF.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=87118255; PubMed=2949326;  
RA Ikuta K., Takami M., Kim C.W., Honjo T., Miyoshi T., Tagaya Y.,  
Kawabe T., Yodoi J.;  
RT "Human lymphocyte Fc receptor for IgE: sequence homology of its  
cloned cDNA with animal lectins.";  
RN Proc. Natl. Acad. Sci. U.S.A. 84:819-823(1987).  
[2]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=87051737; PubMed=2877743;  
RA Kikutani H., Inui S., Sato R., Barsumian E.L., Owaki H.,  
Yanaseki K., Kaisho T., Uchibayashi N., Hardy R.R., Hirano T.,  
Tsunawawa S., Sakiyama F., Suemura M., Kishimoto T.;  
RT "Molecular structure of human lymphocyte receptor for immunoglobulin  
E.";  
RN Cell 47:657-665(1986).  
[3]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=87218454; PubMed=3034567;  
RA Luedin C., Hofstetter H., Sarfati M., Levy C.A., Suter U., Alaimo D.,  
Kilchherr E., Frost H., Delespesse G.;  
RT "Cloning and expression of the cDNA coding for a human lymphocyte IgE  
receptor.";  
RN EMBO J. 6:109-114(1987).  
[4]  
RN PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
RX MEDLINE=93038513; PubMed=1417742;  
RA Rose K., Turcatti G., Graber P., Pochon S., Regamey P.-O.,  
Jansen K.U., Magnenat E., Aubonne N., Bonnefoy J.-Y.;  
RT "Partial characterization of natural and recombinant human soluble  
CD23.";  
RN Biochem. J. 286:819-824(1992).  
[5]  
RN ALTERNATIVE SPLICING.  
RX MEDLINE=89028672; PubMed=2972386;  
RA Yokota A., Kikutani H., Tanaka T., Sato R., Barsumian E.L.,  
Suemura M., Kishimoto T.;  
RT "Two species of human Fc epsilon receptor II (Fc epsilon RII/CD23):  
tissue-specific and IL-4-specific regulation of gene expression.";  
RN Cell 55:611-618(1988).  
[6]  
RN 3D-STRUCTURE MODELING OF LECTIN DOMAIN.  
RP MEDLINE=94191542; PubMed=8142907;  
RA Padlan E.A., Helm B.A.;  
RT "Modeling of the lectin-homology domains of the human and murine low-  
affinity Fc epsilon receptor (Fc epsilon RII/CD23).";  
RN Receptor 3:325-341(1993).  
[7]  
RN 3D-STRUCTURE MODELING OF 173-285.  
RP MEDLINE=96276216; PubMed=8745401;  
RA Bajorath J., Aruffo A.;

RT "Structure-based modeling of the ligand binding domain of the human  
 RT cell surface receptor CD23 and comparison of two independently  
 RL derived molecular models.";  
 CC Protein Sci. 5:240-247(1996).

CC -!- FUNCTION: This receptor has essential roles in the regulation of  
 CC IgE production and in the differentiation of B-cells (it is a B-  
 CC cell-specific antigen).

CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS AS A  
 CC SOLUBLE EXCRETED FORM.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=A;

CC IsoId=P06734-1; Sequence=Displayed;

CC Name=B;

CC IsoId=P06734-2; Sequence=VSP\_003057;

CC -!- PTM: N- and O-glycosylated.

CC -!- MISCELLANEOUS: There are two kinds of Fc receptors for IgE, which  
 CC differ in both structure and function: high affinity receptors on  
 CC basophils and mast cells and low affinity receptors on lymphocytes  
 CC and monocytes.

CC -!- SIMILARITY: Contains 1 C-type lectin family domain.

CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD23 entry; htm".

CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd23.htm".

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M15059; AAA52434.1; -;

DR EMBL; M14766; AAA52435.1; -;

DR EMBL; X04772; CAA28465.1; -;

DR EMBL; M23562; AAA52433.1; -;

DR PIR; A26067; LNHER.

DR PDB; 1HLI; 31-JAN-94.

DR PDB; 1KJE; 03-APR-96.

DR Genew; HGNC:3612; FCER2.

DR MIM; 151445; -;

DR GO; GO:0005887; C: integral to plasma membrane; TAS.

DR GO; GO:0005178; F: integrin binding; TAS.

DR InterPro; IPR002353; Antifreeze1.

DR InterPro; IPR001304; Lectin\_C.

DR Pfam; PF00059; lectin\_c\_1.

DR PRINTS; PR00356; ANTIFREEZE1.

DR SMART; SM00034; CLECT; 1.

DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.

DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.

DR Receptor; Antigen; IgE-binding protein; Repeat; Signal-anchor;

DR Transmembrane; Lectin; Glycoprotein; Alternative splicing;

DR 3D-structure.

FT CHAIN 1 321 MEMBRANE BOUND FORM.

FT CHAIN 150 321 SOLUBLE FORM.

FT DOMAIN 1 21 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 22 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)

FT (POTENTIAL).

FT DOMAIN 48 321 EXTRACELLULAR (POTENTIAL).

FT SITE 162 284 C-TYPE LECTIN (LONG FORM).

FT REPEAT 149 150 CLEAVAGE.

FT REPEAT 69 89

FT REPEAT 90 110

FT REPEAT 111 131

FT DISULFID 160 288

FT DISULFID 163 174

FT DISULFID 191 282

FT DISULFID 259 273

FT CARBOHYD 63 63

FT VARSPLIC 1 7

FT BY SIMILARITY.

FT BY SIMILARITY.

FT N-LINKED (GLCNAC...) (POTENTIAL).

FT MEGQYS -> MNPPSQ (in isoform B).

FT /FTID=VSP\_003057.

FT N -> T (IN REF. 3).

FT CONFLICT 269 269

FT STRAND 174 177

FT HELIX 184 193

FT TURN 194 195

FT STRAND 197 198

FT HELIX 204 214

FT TURN 215 216

FT STRAND 219 228

FT TURN 229 230

FT STRAND 231 234

FT TURN 235 236

FT STRAND 239 239

FT STRAND 245 245

FT TURN 247 248

FT TURN 254 255

FT STRAND 259 262

FT TURN 264 265

FT STRAND 268 271

FT TURN 273 274

FT STRAND 281 284

SQ SEQUENCE 321 AA; 36468 MW; F86708C0E6515B87 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 321;

Best Local Similarity 50.0%; Pred. No. 3.6e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEIXYDX 8

Db 230 GEFIVWDG 237

RESULT 96

UMES3 YEAST

ID UMES3 YEAST STANDARD; PRT; 323 AA.

AC P47821;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE RNA polymerase II holoenzyme cyclin-like subunit.

GN UME3 OR SSN8 OR SRB11 OR YNL025C OR N2805.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]\_SEQUENCE FROM N.A.

RC STRAIN=S288c;

RX MEDLINE=95183147; PubMed=7877695;

RA Liao S.-M., Zhang J., Jeffrey D.A., Koleske A.J., Thompson C.M.,

RA Chao D.M., Viljoen M., van Vuuren H.J.J., Young R.A.;

RT "A kinase-cyclin pair in the RNA polymerase II holoenzyme.";

RL Nature 374:193-196(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288c;

RA Smith J.B., Mallory M.J., Strich R.;

RL Submitted (XXX-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=S288c;

RX MEDLINE=95249601; PubMed=7732022;

RA Kuchin S., Yeghiayan P., Carlson M.;

RT "Cyclin-dependent protein kinase and cyclin homologs SSN3 and SSN8

RT contribute to transcriptional control in yeast.";

RL Proc. Natl. Acad. Sci. U.S.A. 92:4006-4010(1995).

RN [4]

RP SEQUENCE FROM N.A.

RC Duesterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D., Hilbert H.,

RA Moestl D.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RA Andre B., Iraqi Houssaini I., Urrestrazu L.A., Vissers S.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: UME5/SRB10 and UME3/SRB11 form a kinase-cyclin pair in

the RNA polymerase II holoenzyme, and are essential for a normal transcriptional response to galactose induction in vivo, and are involved in CTD (carboxy-terminal domain) phosphorylation and this modification has a role in the response to transcriptional regulators in vivo.

-1- SUBCELLULAR LOCATION: Nuclear (Potential).

-1- SIMILARITY: Belongs to the cyclin family. Cyclin C subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U20221; AAA69820.1; -  
DR EMBL; U16248; AAA64270.1; -  
DR EMBL; U20635; AA85714.1; -  
DR EMBL; Z71301; CAA95887.1; -  
DR PIR; S59373; S59373.  
DR GerMOnline; 143032; -  
DR TRANSFAC; T02155; -  
DR SGD; S0004970; SSN8.  
DR GO; GO:0007126; P.meiosis; IGI.  
DR InterPro; IPR006670; Cyclin.  
DR InterPro; IPR006671; Cyclin\_N.  
DR Pfam; PF00134; cyclin; 1.  
DR SMART; SM00385; CYCLIN; 2.  
DR PROSITE; PS00292; CYCLINS; FALSE\_NEG.  
KW Cyclin; Nuclear protein.  
SQ SEQUENCE 323 AA; 37790 MW; 2BA16A3374CCF207 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 323;  
Best Local Similarity 50.0%; Pred. No. 3.6e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXXDX 8  
:||||:|:  
Db 150 PERIPDP 157

RESULT 97

TF2B ARCFU STANDARD; PRT; 326 AA.

AC O28970;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Transcription initiation factor IIB (TFIIB).  
GN TFB OR AFI299.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyripides N.C.,  
RA Fleischmann R.D., Dougherty B.A., Lee N.H., Sutton G.G., Gill S.,  
RA Kirkness E.F., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., McManis J.F., McDonald L., Utterback T.,  
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.";  
RT Nature 390:364-370(1997).

-1- FUNCTION: Stabilizes TBP binding to an archaeal box-A promoter. Also responsible for recruiting RNA polymerase II to the pre-initiation complex (DNA-TBP-TFIIB) (By similarity).

-1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

-1- SIMILARITY: Belongs to the TFIIB family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AE001014; AAB89947.1; -  
DR PIR; B69412; B69412.  
DR HSP; P29095; IAI5.  
DR TIGR; AF1299; -  
DR HAMAP; MF 00383; -; 1.  
DR InterPro; IPR006670; Cyclin.  
DR InterPro; IPR000812; TFIIB euk.  
DR Pfam; PF00382; transcript\_fac2; 2.  
DR PRINTS; PR06685; TIFACTORIIB.  
DR SMART; SM00385; CYCLIN; 2.  
DR PROSITE; PS00782; TFIIB; 2.  
KW Transcription regulation; Repeat; Zinc-finger; Metal-binding; Zinc; Complete proteome.  
FT ZN\_FING 30 52 ZN-RIBBON TFIIB-TYPE.  
FT REPEAT 143 226 1.  
FT REPEAT 237 318 2.  
FT METAL 30 30 ZINC (BY SIMILARITY).  
FT METAL 33 33 ZINC (BY SIMILARITY).  
FT METAL 49 49 ZINC (BY SIMILARITY).  
FT METAL 52 52 ZINC (BY SIMILARITY).  
SQ SEQUENCE 326 AA; 36939 MW; 9015FE14E00A31E1 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 326;  
Best Local Similarity 50.0%; Pred. No. 3.7e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXXDX 8  
:||||:|:  
Db 45 GEFICQDC 52

RESULT 98

RIR2 ASFM2 STANDARD; PRT; 327 AA.

AC P26713;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ribonucleoside-diphosphate reductase small chain (EC 1.17.4.1)  
DE Ribonucleoside reductase.  
OS African swine fever virus (isolate Malawi Lil 20/1) (ASFV).  
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.  
OX NCBI\_TaxID=10500;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91335775; PubMed=1871976;  
RA Boursnell M., Shaw K., Yanez R.J., Vinuela E., Dixon L.;  
RT "The sequences of the ribonucleotide reductase genes from African swine fever virus show considerable homology with those of the orthopoxvirus, vaccinia virus.";  
RT Virology 184:411-416(1991).  
RL CC -1- FUNCTION: Provides the precursors necessary for DNA synthesis.  
CC -1- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized thiorodoxin + H(2)O = ribonucleoside diphosphate + reduced thiorodoxin.  
CC -1- COFACTOR: Binds 2 iron ions per subunit (By similarity).  
CC -1- PATHWAY: DNA replication pathway; first step.  
CC -1- SUBUNIT: Heterodimer of a large and a small chain.  
CC -1- SIMILARITY: Belongs to the ribonucleoside diphosphate reductase



```

CC small chain family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M64728; -, NOT_ANNOTATED_CDS.
CC PIR; B40568; RDVZAS.
CC HSP; P1157; IXSM.
CC InterPro; IPR000358; Ribonucl redctse.
CC Pfam; PF00268; ribonuc red sm; 1.
CC PROSITE; PS00368; RIBOED_SMALL; 1.
CC Oxidoreductase; DNA replication; Metal-binding; Iron.
CC FT METAL 70 70
CC FT METAL 101 101 IRON 1 (BY SIMILARITY).
CC FT METAL 104 104 IRON 1 AND 2 (BY SIMILARITY).
CC FT METAL 164 164 IRON 1 (BY SIMILARITY).
CC FT METAL 198 198 IRON 2 (BY SIMILARITY).
CC FT METAL 201 201 IRON 2 (BY SIMILARITY).
CC FT ACT_SITE 108 108 BY SIMILARITY.
CC SQ SEQUENCE 327 AA; 38966 MW; E78508DB1978F4B0 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 327;
Best Local Similarity 50.0%; Pred.No. 3.7e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 191 NEFISRDE 198
:||||:|

RESULT 99
ASCD_YERPE STANDARD; PRT; 328 AA.
AC P37911;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE CDP-6-deoxy-L-threo-D-glycero-4-hexulose-3-dehydratase (CDP-6-
DE deoxy-delta-3,4-glucoseen reductase) (EC 1.17.1.-) (E3).
GN ASCD OR RFBI OR YPO3116 OR Y1067.
OS Versinia pestis, and
OS Versinia pseudotuberculosis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Versinia.
OX NCBI_TaxID=632, 633;
RN [1]
RN SEQUENCE FROM N.A.
RC SPECIES=Y.pestis; STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Farhili J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Leatherell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
RA Simmonds M., Skelton J., Oyston P.C.F., Quail M.A., Rutherford K.,
RA "Genome sequence of Versinia pestis, the causative agent of plague."
RT Nature 413:523-527(2001).
RN [2]
RN SEQUENCE FROM N.A.
RC SPECIES=Y.pestis; STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.B., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Versinia pestis KIM.";
```

```

RU J. Bacteriol. 184:4601-4611(2002).
RN [3]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC SPECIES=Y.pseudotuberculosis; STRAIN=V;
RX MEDLINE=94117382; PubMed=8288541;
RA Lo S.F., Miller V.P., Lei Y., Thorson J.S., Liu H.-W., Schottel J.L.;
RT "CDP-6-deoxy-delta 3,4-glucoseen reductase from Versinia
RT pseudotuberculosis: enzyme purification and characterization of the
RT cloned gene."
RL J. Bacteriol. 176:460-468(1994).
RN [4]
RN SEQUENCE FROM N.A.
RC SPECIES=Y.pseudotuberculosis; STRAIN=M85 / Serotype IIA;
RX MEDLINE=93186709; PubMed=8444803;
RA Kessler A.C., Haase A., Reeves P.R.;
RT "Molecular analysis of the 3,6-dideoxyhexose pathway genes of
RT Versinia pseudotuberculosis serogroup IIA."
RL J. Bacteriol. 175:1412-1422(1993).
RN [5]
RN SEQUENCE FROM N.A.
RC SPECIES=Y.pseudotuberculosis; STRAIN=Serotype VA;
RX MEDLINE=94350832; PubMed=8071227;
RA Thorson J.S., Lo S.F., Ploux O., He X., Liu H.-W.;
RT "Studies of the biosynthesis of 3,6-dideoxyhexoses: molecular cloning
RT and characterization of the asc (ascarylose) region from Versinia
RT pseudotuberculosis serogroup VA."
RL J. Bacteriol. 176:5483-5493(1994).
RN [6]
RN CHARACTERIZATION.
RC SPECIES=Y.pseudotuberculosis;
RX MEDLINE=97121283; PubMed=8961949;
RA Johnson D.A., Cassner G.T., Bandarian V., Ruzicka F.J., Ballou D.P.,
RA Reed G.H., Liu H.-W.;
RT "Kinetic characterization of an organic radical in the ascarylose
RT biosynthetic pathway."
RL Biochemistry 35:15846-15856(1996).
RN [7]
RN CHARACTERIZATION.
RC SPECIES=Y.pseudotuberculosis;
RX MEDLINE=96264885; PubMed=8672475;
RA Cassner G.T., Johnson D.A., Liu H.-W., Ballou D.P.;
RT "Kinetics of the reductive half-reaction of the iron-sulfur
RT flavoenzyme CDP-6-deoxy-L-threo-D-glycero-4-hexulose-3-dehydratase."
RL Biochemistry 35:7752-7761(1996).
CC -!- FUNCTION: PARTICIPATES IN THE CONVERSION OF CDP-6-DEOXY-D-GLYCERO-
CC L-THREO-4-HEXULOSE TO 3,6-DIDEOXY-D-GLYCERO-4-HEXULOSE
CC TOGETHER WITH CDP-6-DEOXY-D-GLYCERO-L-THREO-4-HEXULOSE-3-DEHYDRASE
CC (E1) IN TWO CONSECUTIVE STEPS. THE DETAILED MECHANISM OF E3 IS NOT
CC YET RESOLVED.
CC -!- COFACTOR: NADH.
CC -!- PATHWAY: Biosynthesis of CDP-ascarylose, a dideoxyhexose from the
CC cell wall lipopolysaccharide.
CC -!- SUBUNIT: Monomer.
CC -!- SIMILARITY: In the N-terminal section; belongs to the 2Fe2S plant-
CC type ferredoxin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ414155; CAC92352.1; -
CC EMBL; AE013710; AAM84648.1; -
CC EMBL; I25594; AAA16760.1; -
CC EMBL; AF461770; AAB49398.1; -
CC EMBL; L33181; AAA88698.1; -
CC EMBL; S72887; AAB31754.1; -
CC PIR; A36952; A36952.
CC PIR; AE0378; AE0378.
```

```
DR HSSP; P23486; 1QFU.
DR InterPro; IPR006038; 2Fe2S fd BS.
DR InterPro; IPR008333; FAD_binding_6.
DR InterPro; IPR001041; Ferredoxin.
DR InterPro; IPR001709; FPN_cyt_redctse.
DR InterPro; IPR001433; Oxrad_FAD/NAD(P).
DR InterPro; IPR001221; Phe_hydroxylase.
DR Pfam; PF00970; FAD_binding_6; 1.
DR Pfam; PF00111; fer2; 1.
DR Pfam; PF00175; NAD_binding_1; 1.
DR PRINTS; PR00371; PFNCR.
DR PRINTS; PR00410; PHEVDYDXLASE.
DR PROSITE; PS00197; 2PE2S_FERREDOXIN; 1.
KW Oxidoreductase; Electron transport; Metal-binding; Iron-sulfur; Iron;
KW 2Fe-2S; NAD; Complete proteome.
FT INIT MET 0
FT DOMAIN 1 89 FERREDOXIN-LIKE.
FT METAL 36 36 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 41 41 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 44 44 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 74 74 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 328 AA; 35915 MW; 8D03600925CB4012 CRC64;

Query Match 100.0%; Score 25; DB 1; Length 328;
Best Local Similarity 50.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 107 IEFIGEDI 114

RESULT 100
LEU3 METH
ID LEU3 METH STANDARD; PRT; 329 AA.
AC O2741;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3-isopropylmalate dehydrogenase (EC 1.1.1.85) (Beta-IPM dehydrogenase)
DE (IMDH) (3-IPM-DH).
GN LEUB OR MTH1388.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwni N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- FUNCTION: Catalyzes the oxidation of 3-carboxy-2-hydroxy-4-
CC methylpentanoate (3-isopropylmalate) to 3-carboxy-4-methyl-2-
CC oxopentanoate. The product decarboxylates to 4-methyl-2-
CC oxopentanoate.
CC -!- CATALYTIC ACTIVITY: 3-carboxy-2-hydroxy-4-methylpentanoate +
CC NAD(+) = 3-carboxy-4-methyl-2-oxopentanoate + NADH.
CC -!- PATHWAY: Leucine biosynthesis; third step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the isocitrate and isopropylmalate
CC dehydrogenases family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2004, 01:28:37 ; Search time 80 Seconds  
(without alignments)  
31.552 Million cell updates/sec

Title: US-09-660-302E-1  
Perfect score: 25  
Sequence: 1 XEFIXDX 8

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 300 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	32	2	O05663
2	25	100.0	39	11	Q9QXA9
3	25	100.0	44	11	Q9QXC5
4	25	100.0	50	16	Q8EX90
5	25	100.0	54	2	Q9RIK0
6	25	100.0	56	9	Q9AZH6
7	25	100.0	57	2	Q8L2B1
8	25	100.0	65	16	Q8XUJ9
9	25	100.0	67	16	Q8JVM2
10	25	100.0	68	7	Q98248
11	25	100.0	68	7	O19347
12	25	100.0	68	16	Q97N33
13	25	100.0	68	16	Q8XPG0
14	25	100.0	69	2	Q9F572
15	25	100.0	71	16	Q8RHA8
16	25	100.0	74	7	Q31516
17	25	100.0	74	7	Q31516
18	25	100.0	74	7	Q31516
19	25	100.0	74	7	Q31516
20	25	100.0	74	7	Q31516
21	25	100.0	74	7	Q31516
22	25	100.0	74	7	Q31516
23	25	100.0	74	7	Q31516
24	25	100.0	74	7	Q31516
25	25	100.0	74	7	Q31516
26	25	100.0	74	7	Q31516
27	25	100.0	74	7	Q31516
28	25	100.0	74	7	Q31516
29	25	100.0	74	7	Q31516
30	25	100.0	74	7	Q31516
31	25	100.0	74	7	Q31516
32	25	100.0	74	7	Q31516
33	25	100.0	74	7	Q31516
34	25	100.0	74	7	Q31516
35	25	100.0	74	7	Q31516
36	25	100.0	74	7	Q31516
37	25	100.0	74	7	Q31516
38	25	100.0	74	7	Q31516
39	25	100.0	74	7	Q31516
40	25	100.0	74	7	Q31516
41	25	100.0	74	7	Q31516
42	25	100.0	74	7	Q31516
43	25	100.0	74	7	Q31516
44	25	100.0	74	7	Q31516
45	25	100.0	74	7	Q31516
46	25	100.0	74	7	Q31516
47	25	100.0	74	7	Q31516
48	25	100.0	74	7	Q31516
49	25	100.0	74	7	Q31516
50	25	100.0	74	7	Q31516
51	25	100.0	74	7	Q31516
52	25	100.0	74	7	Q31516
53	25	100.0	74	7	Q31516
54	25	100.0	74	7	Q31516
55	25	100.0	74	7	Q31516
56	25	100.0	74	7	Q31516
57	25	100.0	74	7	Q31516
58	25	100.0	74	7	Q31516
59	25	100.0	74	7	Q31516
60	25	100.0	74	7	Q31516
61	25	100.0	74	7	Q31516
62	25	100.0	74	7	Q31516
63	25	100.0	74	7	Q31516
64	25	100.0	74	7	Q31516
65	25	100.0	74	7	Q31516
66	25	100.0	74	7	Q31516
67	25	100.0	74	7	Q31516
68	25	100.0	74	7	Q31516
69	25	100.0	74	7	Q31516
70	25	100.0	74	7	Q31516
71	25	100.0	74	7	Q31516
72	25	100.0	74	7	Q31516
73	25	100.0	74	7	Q31516
74	25	100.0	74	7	Q31516
75	25	100.0	74	7	Q31516
76	25	100.0	74	7	Q31516
77	25	100.0	74	7	Q31516
78	25	100.0	74	7	Q31516
79	25	100.0	74	7	Q31516
80	25	100.0	74	7	Q31516
81	25	100.0	74	7	Q31516
82	25	100.0	74	7	Q31516
83	25	100.0	74	7	Q31516
84	25	100.0	74	7	Q31516
85	25	100.0	74	7	Q31516
86	25	100.0	74	7	Q31516
87	25	100.0	74	7	Q31516
88	25	100.0	74	7	Q31516
89	25	100.0	74	7	Q31516

Q8V3K6 swinepox vi  
Q91MS0 lumpy skin  
Q8JTS8 lumpy skin  
Q84Q13 chlamydomon  
Q97EC0 clostridium  
Q9FW17 oryza sativ  
Q7XHE3 oryza sativ  
Q88WY2 lactobacill  
Q81W38 bacillus an  
Q812V8 bacillus ce  
Q95HS2 peromyscus  
Q8U1A6 pyrococcus  
Q89179 alcaligenes  
Q51940 burkholderi  
Q8PJA6 xanthomonas  
Q8ET09 oceanobacil  
Q8GJV0 oryctolagus  
Q7T1F7 gallus gall  
Q93EH4 helicobacte  
Q24060 drosophila  
Q81417 caenorhabdi  
Q8D9P7 vibrio vuln  
Q8D9R7 vibrio vuln  
Q24077 drosophila  
Q7WTJ7 acinetobact  
Q863N9 arctonyx co  
Q92528 carnation l  
Q8J1L1 carnation l  
Q81J6 pseudomonas  
Q82973 bacillus sp  
Q863Q6 lontra cana  
Q863P4 eira barbar  
Q8C3V0 mus musculu  
Q863Q3 lutra lutra  
Q863Q2 lutra macul  
Q863Q9 aonyx capen  
Q863Q8 amblyonyx ci  
Q863Q7 enhydra lut  
Q863Q5 lontra feli  
Q863Q4 lontra long  
Q863Q1 pteronura b  
Q863Q0 mustela erm  
Q863P9 mustela fire  
Q863P8 mustela vis  
Q863P7 martes amer  
Q863P6 martes penn  
Q863P5 gulo gulo  
Q863P3 galictis vi  
Q863P2 ictonyx str  
Q863P1 meles meles  
Q863P0 taxidea tax  
Q863N8 melogale mo  
Q863N7 bassariscus  
Q863N6 procyon lot  
Q863N5 streptococc  
Q863N4 raietonia s  
Q863N3 helicobacte  
Q83UP9 gamma-prote  
Q8H82 oryza sativ  
Q41901 arabidopsis  
Q9M96 arabidopsis  
Q8G2V8 brucella su  
Q9AZU2 bacterioph  
Q8CF8 lactococcus  
Q81T0 arabidopsis  
Q8XCO1 escherichia  
Q7VBN6 prochloroco  
Q81V13 bacillus ce  
Q81W3 bacillus an  
Q9KMD8 vibrio chol  
Q89GN0 streptomyce  
Q8EX15 mycoplasma  
Q97E59 clostridium

90	25	100.0	123	16	QAVEY2	Q8vey2 bruceella me	163	25	100.0	166	13	Q918D3	Q918d3 gallus gall
91	25	100.0	123	16	Q8EH18	Q8eh18 shewanella	164	25	100.0	167	16	Q8YGU0	Q8ygu0 bruceella me
92	25	100.0	124	2	Q93F05	Q93fu5 cowdria rum	165	25	100.0	167	16	Q8G128	Q8g128 bruceella su
93	25	100.0	124	10	Q8L3Q2	Q8l3q2 oryza sativ	166	25	100.0	168	13	Q918D4	Q918d4 gallus gall
94	25	100.0	124	16	Q9CLP6	Q9clp6 pasteurella	167	25	100.0	169	16	Q8D2H1	Q8d2h1 wigglewort
95	25	100.0	127	9	Q8S9F7	Q8s9f7 pseudomonas	168	25	100.0	170	8	Q8BAF6	Q8baf6 ranunculus
96	25	100.0	129	12	Q89445	Q89445 african swi	169	25	100.0	170	16	Q97IG5	Q97ig5 clostridium
97	25	100.0	129	12	Q9DHS1	Q9dhs1 yaba-like d	170	25	100.0	171	10	Q9FHW7	Q9fhw7 arbidopsis
98	25	100.0	130	7	Q8E225	Q8e225 hypogeomys	171	25	100.0	171	16	Q84305	Q84305 chlamydia t
99	25	100.0	130	10	Q8GUR8	Q8gur8 pisum sativ	172	25	100.0	171	16	Q9K604	Q9k604 bacillus h
100	25	100.0	130	11	Q99MY7	Q99my7 meriones un	173	25	100.0	172	10	Q6S283	Q6s283 arbidopsis
101	25	100.0	130	12	Q8JTV9	Q8jtv9 lumpy skin	174	25	100.0	173	16	Q8A6U2	Q8a6u2 bacteroides
102	25	100.0	130	12	Q91MW6	Q91mw6 lumpy skin	175	25	100.0	173	17	Q8TLM2	Q8tlm2 methanosarc
103	25	100.0	131	16	Q8FR36	Q8fr36 xanthomonas	176	25	100.0	173	17	Q8QO50	Q8qo50 methanosarc
104	25	100.0	132	6	Q863K5	Q863k5 oryctolagus	177	25	100.0	174	2	Q93HV7	Q93hy7 enterococcu
105	25	100.0	132	13	Q90Z73	Q90z73 oncorhynch	178	25	100.0	174	16	Q8Y2Y2	Q8y2y2 ralstonia s
106	25	100.0	132	13	Q8AY18	Q8ay18 cyprinus ca	179	25	100.0	175	9	Q7Y544	Q7y544 bacterioph
107	25	100.0	133	5	Q86HJ5	Q86hj5 dictyosteli	180	25	100.0	176	16	Q97IZ1	Q97iz1 clostridium
108	25	100.0	133	17	Q8PUF3	Q8puf3 methanosarc	181	25	100.0	176	16	Q8A2A0	Q8a2a0 bacteroides
109	25	100.0	133	16	Q8YAH5	Q8yah5 listeria mo	182	25	100.0	176	16	Q7V2H7	Q7v2h7 prochloroco
110	25	100.0	136	16	Q8L1O55	Q8l1u55 bacillus an	183	25	100.0	176	16	Q7X30	Q7x30 sulfolobus
111	25	100.0	137	10	Q9LUP94	Q9lp94 arbidopsis	184	25	100.0	177	16	Q97G88	Q97g88 clostridium
112	25	100.0	137	16	Q98L34	Q98l34 rhizobium l	185	25	100.0	178	2	Q8KM04	Q8km04 photorhabdu
113	25	100.0	137	16	Q9L4R9	Q9l4r9 bacillus ce	186	25	100.0	178	2	Q8KM07	Q8km07 photorhabdu
114	25	100.0	139	5	Q7YX35	Q7yy35 cryptospori	187	25	100.0	178	2	Q8P445	Q8p445 xanthomonas
115	25	100.0	140	2	Q8ROF9	Q8rof9 leptospira	188	25	100.0	179	16	Q8P445	Q8p445 xanthomonas
116	25	100.0	140	2	Q7S0Z6	Q7s0z6 escherichia	189	25	100.0	180	16	Q983M6	Q983m6 rhizobium l
117	25	100.0	140	2	Q9ANS8	Q9ans8 mycobacteri	189	25	100.0	180	16	Q983M6	Q983m6 rhizobium l
118	25	100.0	141	9	Q9G092	Q9g092 leptospira	190	25	100.0	181	10	Q91ZL2	Q91z12 arbidopsis
119	25	100.0	141	4	Q9NR14	Q9nr14 homo sapien	191	25	100.0	181	17	Q9YBU6	Q9ybu6 aeropyrum p
120	25	100.0	141	4	Q9H5V3	Q9h5v3 homo sapien	192	25	100.0	181	17	Q5A155	Q5a155 methanococ
121	25	100.0	142	10	Q941A6	Q941a6 arbidopsis	193	25	100.0	182	5	Q8T965	Q8t965 drosophila
122	25	100.0	143	10	Q9LE10	Q9le10 arbidopsis	194	25	100.0	183	16	Q8EW3	Q8ew3 mycoplasma
123	25	100.0	144	12	Q66047	Q66047 canine herp	195	25	100.0	184	16	Q8EPFQ1	Q8epfq1 xanthomonas
124	25	100.0	144	12	Q66071	Q66071 canine herp	196	25	100.0	184	16	Q8EFE6	Q8efe6 shewanella
125	25	100.0	145	9	Q9MBX8	Q9mbx8 streptococ	197	25	100.0	185	8	Q9G923	Q9g923 ochromonas
126	25	100.0	145	16	Q98BM6	Q98bm6 rhizobium l	198	25	100.0	186	2	Q8S650	Q8s650 streptococ
127	25	100.0	145	16	Q92SM7	Q92sm7 rhizobium m	199	25	100.0	186	2	Q53147	Q53147 rhodococcus
128	25	100.0	146	10	Q9LWD2	Q9lwd2 oryza sativ	200	25	100.0	186	5	Q7YYF2	Q7yyf2 cryptospori
129	25	100.0	146	12	Q9YUHI	Q9yuh1 rotavirus b	201	25	100.0	186	13	Q91961	Q91961 gallus gall
130	25	100.0	147	9	Q8SD36	Q8sd36 pseudomonas	202	25	100.0	187	16	Q26678	Q26678 methanobact
131	25	100.0	148	2	Q96980	Q96980 rhodobacter	203	25	100.0	187	16	Q97JA8	Q97ja8 clostridium
132	25	100.0	148	5	Q9U4X7	Q9u4x7 strongyloid	204	25	100.0	188	2	Q8VLS2	Q8vls2 versinia en
133	25	100.0	148	5	Q9VZL1	Q9vzl1 drosophila	205	25	100.0	188	2	Q8VLS2	Q8vls2 versinia en
134	25	100.0	148	10	Q7XS46	Q7xs46 oryza sativ	206	25	100.0	189	16	Q92JH4	Q92jhm4 rickettsia
135	25	100.0	149	16	Q97G33	Q97g33 clostridium	207	25	100.0	189	16	Q8LJY3	Q8lly3 bacillus an
136	25	100.0	151	16	Q9PN39	Q9pn39 campylobact	208	25	100.0	190	16	Q8LJY3	Q8lly3 bacillus an
137	25	100.0	152	6	Q8WNM2	Q8wnm2 gorilla gor	209	25	100.0	191	8	Q8WJ98	Q8wjl98 anogramma g
138	25	100.0	154	2	Q06462	Q06462 staphylococ	210	25	100.0	191	16	Q8KCL0	Q8kcl0 chlorobium
139	25	100.0	154	2	Q9ZANO	Q9zan0 staphylococ	211	25	100.0	192	11	Q8BIM9	Q8bim9 mus musculu
140	25	100.0	154	16	Q9A9W7	Q9a9w7 caulobacter	212	25	100.0	193	11	Q8BIM9	Q8bim9 mus musculu
141	25	100.0	155	16	Q97MM9	Q97mm9 clostridium	213	25	100.0	194	10	Q22891	Q22891 arbidopsis
142	25	100.0	156	9	Q2ZWZ5	Q9zwz5 mycobacteri	214	25	100.0	196	2	Q9AME2	Q9ame2 escherichia
143	25	100.0	156	9	Q2ZWZ5	Q9zwz5 mycobacteri	215	25	100.0	196	16	Q8X611	Q8x611 escherichia
144	25	100.0	156	12	Q89235	Q89235 white clove	216	25	100.0	196	16	Q8X611	Q8x611 escherichia
145	25	100.0	156	16	Q93DB5	Q93db5 streptococ	217	25	100.0	196	16	Q8FKL2	Q8fkl2 escherichia
146	25	100.0	157	4	Q8WY66	Q8wy66 homo sapien	218	25	100.0	199	10	Q65035	Q65035 oryza sativ
147	25	100.0	157	11	Q8CC13	Q8cci3 mus musculu	219	25	100.0	199	16	Q8KAY7	Q8kay7 chlorobium
148	25	100.0	157	11	Q8CC13	Q8cci3 mus musculu	220	25	100.0	200	5	Q8I1I8	Q8i1i8 plasmodium
149	25	100.0	157	16	Q74394	Q74394 synechocyst	221	25	100.0	201	2	Q7WVX8	Q7wvx8 anaplasma s
150	25	100.0	158	12	Q9EN67	Q9en67 spodoptera	222	25	100.0	201	2	Q7WVX8	Q7wvx8 anaplasma s
151	25	100.0	160	16	Q9KPI0	Q9kpi0 drosophila	223	25	100.0	201	10	Q7Y1Z7	Q7y1z7 arbidopsis
152	25	100.0	160	16	Q928E7	Q928e7 vibrio chol	224	25	100.0	201	16	Q8CPX4	Q8cp4 staphylococ
153	25	100.0	160	16	Q928E7	Q928e7 listeria in	225	25	100.0	202	3	Q9P7W1	Q9p7wl schizosach
154	25	100.0	160	16	Q813C7	Q813c7 bacillus ce	226	25	100.0	202	16	Q92DQ9	Q92dd9 listeria in
155	25	100.0	161	2	Q30617	Q30617 myxococcus	227	25	100.0	202	16	Q8Y8Y2	Q8y8y2 listeria mo
156	25	100.0	161	9	Q8W5V8	Q8w5v8 bacterioph	228	25	100.0	202	16	Q8Y8Y2	Q8y8y2 listeria mo
157	25	100.0	161	13	Q918F2	Q918f2 esox lucius	229	25	100.0	202	16	Q8Y8Y2	Q8y8y2 listeria mo
158	25	100.0	163	8	Q7YFV1	Q7yfv1 neurospora	230	25	100.0	202	16	Q8Y8Y2	Q8y8y2 listeria mo
159	25	100.0	163	16	Q8KB07	Q8kb07 bacillus ha	231	25	100.0	202	16	Q8Y8Y2	Q8y8y2 listeria mo
160	25	100.0	164	2	Q8GJ7D	Q8gjd7 campylobact	232	25	100.0	202	16	Q8Y8Y2	Q8y8y2 listeria mo
161	25	100.0	164	5	Q818Z9	Q818z9 euplores cr	233	25	100.0	203	16	Q8Y8Y2	Q8y8y2 listeria mo
162	25	100.0	164	5	Q8H5C3	Q8h5c3 oryza sativ	234	25	100.0	203	16	Q8Y8Y2	Q8y8y2 listeria mo
163	25	100.0	164	16	Q9PND0	Q9pnd0 campylobact	235	25	100.0	204	5	Q9U8Q6	Q9u8q6 pseudomonas
164	25	100.0	164	17	Q8U1K5	Q8u1k5 pyrococcus	235	25	100.0	204	5	Q9U8Q6	Q9u8q6 pseudomonas
165	25	100.0	164	17	Q8U1K5	Q8u1k5 pyrococcus	235	25	100.0	204	5	Q9U8Q6	Q9u8q6 pseudomonas

236 Q8u9K8 agrobacteri 25 100.0 205 16 Q8U9K8  
 237 Q8Y2E2 ralstonia s 25 100.0 206 16 Q8Y2E2  
 238 Q8M47 rhizobium l 25 100.0 207 16 Q8M47  
 239 Q92GW7 rickettsia 25 100.0 207 16 Q92GW7  
 240 Q8B25 pseudomonas 25 100.0 207 16 Q8B25  
 241 Q87A23 xyella fas 25 100.0 207 16 Q87A23  
 242 Q9ufK1 homo sapien 25 100.0 208 4 Q9ufK1  
 243 Q9pa77 xyella fas 25 100.0 208 16 Q9PA77  
 244 Q92WD5 rhizobium m 25 100.0 208 16 Q92WD5  
 245 Q92CA9 rickettsia 25 100.0 209 16 Q92CA9  
 246 Q92G07 rickettsia 25 100.0 209 16 Q92G07  
 247 Q8PC62 xanthomonas 25 100.0 209 16 Q8PC62  
 248 Q7WMD3 bordetella 25 100.0 210 16 Q7WMD3  
 249 Q7W8R6 bordetella 25 100.0 210 16 Q7W8R6  
 250 Q7VZ56 bordetella 25 100.0 210 16 Q7VZ56  
 251 Q7WYT0 rhizobium l 25 100.0 211 2 Q7WYT0  
 252 Q85SK5 mycobacteri 25 100.0 211 9 Q85SK5  
 253 Q85J53 arabidopsis 25 100.0 211 10 Q85J53  
 254 Q8H9D3 solanum tub 25 100.0 211 10 Q8H9D3  
 255 Q8SST3 arabidopsis 25 100.0 211 10 Q8SST3  
 256 Q8PNT9 xanthomonas 25 100.0 211 16 Q8PNT9  
 257 Q7WR85 bordetella 25 100.0 211 16 Q7WR85  
 258 Q7W2B7 bordetella 25 100.0 211 16 Q7W2B7  
 259 Q7VT94 bordetella 25 100.0 211 16 Q7VT94  
 260 Q914X9 pseudomonas 25 100.0 212 16 Q914X9  
 261 Q89NR4 bradyrhizob 25 100.0 212 16 Q89NR4  
 262 Q8W5V5 bacterioph 25 100.0 213 9 Q8W5V5  
 263 Q9PD90 xyella fas 25 100.0 213 16 Q9PD90  
 264 Q88HV7 rhizobium l 25 100.0 213 16 Q88HV7  
 265 Q8CV08 oceanobacil 25 100.0 213 16 Q8CV08  
 266 Q87DH9 xyella fas 25 100.0 213 16 Q87DH9  
 267 Q8WVE0 homo sapien 25 100.0 214 4 Q8WVE0  
 268 Q9CY45 mus musculu 25 100.0 214 11 Q9CY45  
 269 Q51727 borrelia bu 25 100.0 214 16 Q51727  
 270 Q883F6 pseudomonas 25 100.0 214 16 Q883F6  
 271 Q81S18 bacillus an 25 100.0 214 16 Q81S18  
 272 Q29541 archaeoglob 25 100.0 214 17 Q29541  
 273 Q9R2T6 streptococ 25 100.0 215 2 Q9R2T6  
 274 Q9AT50 fucus vesic 25 100.0 215 10 Q9AT50  
 275 Q93P81 microscilla 25 100.0 216 2 Q93P81  
 276 Q7V0T3 prochloroco 25 100.0 216 16 Q7V0T3  
 277 Q88PX7 pseudomonas 25 100.0 217 16 Q88PX7  
 278 Q7V1W1 prochloroco 25 100.0 217 16 Q7V1W1  
 279 Q01594 caenorhabdi 25 100.0 218 5 Q01594  
 280 Q39127 arabidopsis 25 100.0 218 10 Q39127  
 281 Q8ENG2 oceanobacil 25 100.0 218 16 Q8ENG2  
 282 Q880F1 sulfolobus 25 100.0 218 17 Q880F1  
 283 Q81M90 bacillus an 25 100.0 219 16 Q81M90  
 284 Q7WHH3 bordetella 25 100.0 219 16 Q7WHH3  
 285 Q7W994 xyella fas 25 100.0 219 16 Q7W994  
 286 Q7VYQ1 bordetella 25 100.0 219 17 Q7VYQ1  
 287 Q978A6 thermoplas 25 100.0 219 17 Q978A6  
 288 Q81P71 bacillus an 25 100.0 220 16 Q81P71  
 289 Q81C51 bacillus ce 25 100.0 220 16 Q81C51  
 290 Q818V5 bacillus ce 25 100.0 220 16 Q818V5  
 291 Q97KD2 clostridium 25 100.0 221 16 Q97KD2  
 292 Q7VCY6 prochloroco 25 100.0 221 16 Q7VCY6  
 293 Q8GJC7 campylobact 25 100.0 222 2 Q8GJC7  
 294 Q9KVM8 vibrio chol 25 100.0 222 16 Q9KVM8  
 295 Q8XAE0 escherichia 25 100.0 222 16 Q8XAE0  
 296 Q9MQZ1 laqueus rub 25 100.0 223 8 Q9MQZ1  
 297 Q8Y5A3 listeria mo 25 100.0 223 16 Q8Y5A3  
 298 Q91V63 arabidopsis 25 100.0 224 10 Q91V63  
 299 Q8GSH7 oryza sativ 25 100.0 224 10 Q8GSH7  
 300 Q9CQR0 mus musculu 25 100.0 224 11 Q9CQR0

## ALIGNMENTS

RESULT 1

ID Q05663

PRELIMINARY; PRT; 32 AA.

AC Q05663;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Dihydro-oxotase (Fragment).  
 OS Lactobacillus leichmannii.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillus.  
 OX NCBI\_TaxID=28039;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM20076;  
 RX MEDLINE=96285745; PubMed=8725005;  
 RA Becker J., Brendel M.;  
 RT "Molecular cloning and characterization of the pyrB gene of  
 Lactobacillus leichmannii encoding aspartate transcarbamylase.";  
 RL Biochimie 78:3-3(1996).  
 DR EMBL; X84262; CAA59022.1; -.  
 DR InterPro; IPR005847; Pept\_M38\_regn.  
 DR ProDom; PD000518; Urease; 1.  
 FT NON\_TER 32  
 SQ SEQUENCE 32 AA; 3505 MW; 85797BB62D33540E CRC64;

Query Match 100.0%; Score 25; DB 2; Length 32;

Best Local Similarity 50.0%; Pred. No. 2.9e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8

Db 14 GEFIKEDV 21

## RESULT 2

ID Q9QXA9 PRELIMINARY; PRT; 39 AA.  
 AC Q9QXA9;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Fructose-1,6-bisphosphatase (EC 3.1.3.11) (Fragment).  
 GN FBPI.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=21147931; PubMed=11250076;  
 RA Stein S., Liehr T., Eschrich K.;  
 RT "Characterization of the mouse liver fructose-1,6-bisphosphatase  
 gene.";  
 RL Gene 264:215-224(2001).  
 DR EMBL; AJ245389; CAB65268.1; -.  
 DR HSP; P00636; 1CNO.  
 DR MGD; MGI:95492; Fbp1.  
 DR GO; GO:0042132; F:fructose-bisphosphatase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0042578; F:phosphoric ester hydrolase activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR000146; In\_FB\_phptase.  
 DR Pfam; PF00316; FBPase; 1.  
 DR ProDom; PD001491; In\_FB\_phptase; 1.  
 KW Hydrolase.  
 FT NON\_TER 39  
 SQ SEQUENCE 39 AA; 4513 MW; 4B5CE6D9E980B0FB CRC64;

Query Match 100.0%; Score 25; DB 11; Length 39;

Best Local Similarity 50.0%; Pred. No. 3.6e+02;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8

Db	3 GEFIMVDR 10	:   : :	1 XEFIXDX 8	:   : :	6 SEFIKNDI 13
RESULT 3					
Q9QXC5	PRELIMINARY;	PRT;	44 AA.		
ID	Q9QXC5				
AC	Q9QXC5				
DT	01-MAY-2000 (TrEMBLrel. 13, Created)				
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Fructose-1,6-bisphosphatase (EC 3.1.3.11) (Fragment).				
GN	FBP1 OR FBPASE-1.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Liver;				
RX	MEDLINE=21147931; PubMed=11250076;				
RA	Stein S., Liehr T., Eschrich K.;				
RT	"Characterization of the mouse liver fructose-1,6-bisphosphatase				
RT	gene.";				
RL	Gene 264:215-224(2001).				
DR	EMBL; AJ242919; CAB65255.1; -.				
DR	HSP; P00636; ILCNQ.				
DR	MGI; MGI:95492; Fbpl.				
DR	GO; GO:042132; F:fructose-bisphosphatase activity; IEA.				
DR	GO; GO:0016787; F:hydrolase activity; IEA.				
DR	GO; GO:0042578; F:phosphoric ester hydrolase activity; IEA.				
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.				
DR	InterPro; IPR000146; In_FB_phptase.				
DR	Pfam; PF00316; FBPase; 1.				
DR	ProDom; PD001491; In_FB_phptase; 1.				
KW	Hydrolase.				
FT	NON_TER	1			
FT	NON_TER	44			
FT	NON_TER	44			
SQ	SEQUENCE 44 AA; 5195 MW; 34206AD84191E8D CRC64;				
Query Match	100.0%; Score 25; DB 11; Length 44;				
Best Local Similarity	50.0%; Pred. No. 4.1e+02;				
Matches	4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;				
Qy	1 XEFIXDX 8				
Db	3 GEFIMVDR 8	:   : :			
RESULT 4					
Q8EX90	PRELIMINARY;	PRT;	50 AA.		
ID	Q8EX90				
AC	Q8EX90				
DT	01-MAR-2003 (TrEMBLrel. 23, Created)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)				
DE	Hypothetical protein.				
GN	LB324.				
OS	Leptospira interrogans.				
OC	Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.				
OX	NCBI_TaxID=173;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=56601 / Serogroup Interohaemorrhagiae / Seroovar lai;				
RA	Ren S.;				
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AE011619; AAN51883.1; -.				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 50 AA; 6252 MW; 871A05BF16474779 CRC64;				
Query Match	100.0%; Score 25; DB 16; Length 50;				
Best Local Similarity	50.0%; Pred. No. 4.7e+02;				
Matches	4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;				
Qy	1 XEFIXDX 8				
Db	3 GEFIMVDR 8	:   : :			
RESULT 5					
Q9RIK0	PRELIMINARY;	PRT;	54 AA.		
ID	Q9RIK0				
AC	Q9RIK0				
DT	01-MAY-2000 (TrEMBLrel. 13, Created)				
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	DNA helicase homolog (Fragment).				
OS	Streptococcus pyogenes.				
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
OC	Streptococcus.				
OX	NCBI_TaxID=1314;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CS101;				
RA	Wolschik M., Fodbielski A.;				
RL	Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; U35377; AAF08588.1; -.				
DR	HSP; P03005; IJWE.				
DR	GO; GO:0005524; F:ATP binding; IEA.				
DR	GO; GO:0003678; F:DNA helicase activity; IEA.				
DR	GO; GO:0006260; P:DNA replication; IEA.				
DR	InterPro; IPR007693; DnaB_N.				
DR	Pfam; PF00772; DnaB; 1.				
KW	Helicase.				
FT	NON_TER	1			
FT	NON_TER	54			
FT	NON_TER	54			
SQ	SEQUENCE 54 AA; 6222 MW; E053B5D20762B9DA CRC64;				
Query Match	100.0%; Score 25; DB 2; Length 54;				
Best Local Similarity	50.0%; Pred. No. 5.1e+02;				
Matches	4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;				
Qy	1 XEFIXDX 8				
Db	28 REFIRPDD 35	:   : :			
RESULT 6					
Q9AZH6	PRELIMINARY;	PRT;	56 AA.		
ID	Q9AZH6				
AC	Q9AZH6				
DT	01-JUN-2001 (TrEMBLrel. 17, Created)				
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)				
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)				
DE	ORF5.				
GN	ORF5.				
OS	Bacteriophage b1311.				
OC	Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.				
OX	NCBI_TaxID=151534;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=21113149; PubMed=11160885;				
RA	Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;				
RT	"Analysis of six prophages in Lactococcus lactis I1403: different				
RT	genetic structure of temperate and virulent phage populations.";				
RL	Nucleic Acids Res. 29:644-651(2001).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;				
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF323672; AAK08437.1; -.				
SQ	SEQUENCE 56 AA; 6356 MW; 522F338FF4562DB0 CRC64;				
Query Match	100.0%; Score 25; DB 9; Length 56;				
Best Local Similarity	50.0%; Pred. No. 5.3e+02;				
Matches	4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;				

```

Qy 1 XEPIXDX 8
   :|||::|
Db 2 TERITIDE 9

RESULT 7
Q8L2B1
ID Q8L2B1 PRELIMINARY; PRT; 57 AA.
AC Q8L2B1;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Arsenate reductase.
GN ORF90.
OS Proteus vulgaris.
OG Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Proteus.
OX NCBI_TaxID=585;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=UR-75;
RA Murata T., Hayashi T.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=UR-75;
RX MEDLINE=22024716; PubMed=12029035;
RA Murata T., Ohnishi M., Ara T., Kaneko J., Han C.-G., Li Y.F.,
RA Takashima K., Nojima H., Nakayama K., Kaji A., Kamio Y., Miki T.,
RA Mori H., Ohtsubo E., Terawaki Y., Hayashi T.;
RT "Complete Nucleotide Sequence of Plasmid Rts1: Implications for
RT Evolution of Large Plasmid Genomes.";
RL J. Bacteriol. 184:3194-3202(2002).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=UR-75;
RX MEDLINE=96184644; PubMed=8645296;
RA Tian Q.B., Ohnishi M., Tabuchi A., Terawaki Y.;
RT "A new plasmid-encoded proteic killer gene system: cloning,
RT sequencing, and analyzing hig locus of plasmid Rts1.";
RL Biochem. Biophys. Res. Commun. 220:280-284(1996).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=UR-75;
RX MEDLINE=94358883; PubMed=8078071;
RA Janosi L., Yonemitsu H., Hong H., Kaji A.;
RT "Molecular cloning and expression of a novel hydroxymethylcytosine-
RT specific restriction enzyme (PvuRts1I) modulated by glucosylation of
RT DNA.";
RL J. Mol. Biol. 242:45-61(1994).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=UR-75;
RX MEDLINE=91193219; PubMed=2013575;
RA Mochida S., Tsuchiya H., Mori K., Kaji A.;
RT "Three short fragments of Rts1 DNA are responsible for the
RT temperature-sensitive growth phenotype (tsg) of host bacteria.";
RL J. Bacteriol. 173:2600-2607(1991).
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=UR-75;
RX MEDLINE=88289663; PubMed=2840681;
RA Nozue H., Tsuchiya K., Kamio Y.;
RT "Nucleotide sequence and copy control function of the extension of the
RT inci region (inci-b) of Rts1.";
RL Plasmid 19:46-56(1988).
[7]
RP SEQUENCE FROM N.A.
RC STRAIN=UR-75;
RX MEDLINE=88139175; PubMed=3277947;
RA Tanaka M., Okawa N., Mori K., Suyama Y., Kaji A.;

```

```

RT "Nucleotide sequence of an Rts1 fragment causing temperature-dependent
RT instability.";
RN J. Bacteriol. 170:1175-1182(1988).
[8]
RP SEQUENCE FROM N.A.
RC STRAIN=UR-75;
RX MEDLINE=85234397; PubMed=2989253;
RA Mollet B., Clerget M., Meyer J., Iida S.;
RT "Organization of the Tn6-related kanamycin resistance transposon
RT Tn2680 carrying two copies of IS26 and an IS903 variant, IS903. B.";
RN J. Bacteriol. 163:55-60(1985).
[9]
RP SEQUENCE FROM N.A.
RC STRAIN=UR-75;
RX MEDLINE=84185439; PubMed=6325393;
RA Kamio Y., Tabuchi A., Itoh Y., Katagiri H., Terawaki Y.;
RT "Complete nucleotide sequence of mini-Rts1 and its copy mutant.";
RN J. Bacteriol. 158:307-312(1984).
[10]
RP SEQUENCE FROM N.A.
RC STRAIN=UR-75;
RX MEDLINE=83290717; PubMed=6309744;
RA Kamio Y., Terawaki Y.;
RT "Nucleotide sequence of an incompatibility region of mini-Rts1 that
RT contains five direct repeats.";
RL J. Bacteriol. 155:1185-1191(1983).
[11]
RP SEQUENCE FROM N.A.
RC STRAIN=UR-75;
RX MEDLINE=68393387; PubMed=4876466;
RA Terawaki Y., Kakizawa Y., Takayasu H., Yoshikawa M.;
RT "Temperature sensitivity of cell growth in Escherichia coli associated
RT with the temperature sensitive R(KW) factor.";
RL Nature 219:284-285(1968).
DR EMBL; AP004237; BAB93653.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Plasmid.
SQ SEQUENCE 57 AA; 6584 MW; B2AFFA7D5790A1AF CRC64;
Query Match 100.0%; Score 25; DB 2; Length 57;
Best Local Similarity 50.0%; Pred. No. 5.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Qy 1 XEPIXDX 8
   :|||::|
Db 38 REFIEDG 45

RESULT 8
Q8XUJ9
ID Q8XUJ9 PRELIMINARY; PRT; 65 AA.
AC Q8XUJ9;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein RSC3189.
GN RSC3189 OR RS06087.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Bottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).

```

```

DR EMBL; AL646074; CAD16977.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 65 AA; 7398 MW; C63E96C2631ED973 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 65;
Best Local Similarity 50.0%; Pred. No. 6.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 35 EEFIRSDV 42

RESULT 9
Q9JVM2 PRELIMINARY; PRT; 67 AA.
AC Q9JVM2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein NMA0780.
GN NMA0780.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OC NCBI_TaxID=65699;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Z2491 / Serogroup A / Serotype 4A;
RC MEDLINE=2022556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491";
RL Nature 404:502-506(2000).
DR EMBL; AL162754; CAB84063.1; -.
DR PIR; E81922; E81922.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 67 AA; 8031 MW; 34350E84C0AD0DB8 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 67;
Best Local Similarity 50.0%; Pred. No. 6.4e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 16 YEFIFPDQ 23

RESULT 10
O38248 PRELIMINARY; PRT; 68 AA.
ID O38248;
AC O38248;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE MHC class II DQ-alpha (Fragment).
DQA.
GN DQA.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OC NCBI_TaxID=9796;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Blood;
RC Fraser D.G., Bailey E.;
RA "Extensive Polymorphism at the Horse DQA Locus";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF115325; AAD19975.1; -.

```

---

```

DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR001003; MHC_II_alpha.
DR Pfam; PF00993; MHC_II_alpha; 1.
FT NON TER 1 68
FT NON TER 68 68
SQ SEQUENCE 68 AA; 8004 MW; 8A522AE25C51581 CRC64;

Query Match 100.0%; Score 25; DB 7; Length 68;
Best Local Similarity 50.0%; Pred. No. 6.5e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 40 SEFISFDP 47

RESULT 11
O19347 PRELIMINARY; PRT; 68 AA.
ID O19347;
AC O19347;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE MHC class II DQ-alpha (Fragment).
GN ELA-DQA.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OC NCBI_TaxID=9796;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Blood;
RC MEDLINE=98221123; PubMed=9553156;
RA Fraser D.G., Bailey E.;
RT "Polymorphism and multiple loci for the horse DQA gene.";
RL Immunogenetics 47:487-490(1998).
DR EMBL; U92514; AAC17578.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR001003; MHC_II_alpha.
DR Pfam; PF00993; MHC_II_alpha; 1.
FT NON TER 1 68
FT NON TER 68 68
SQ SEQUENCE 68 AA; 7992 MW; 8A522AF80CC51581 CRC64;

Query Match 100.0%; Score 25; DB 7; Length 68;
Best Local Similarity 50.0%; Pred. No. 6.5e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 40 SEFISFDP 47

RESULT 12
O97N33 PRELIMINARY; PRT; 68 AA.
ID O97N33;
AC O97N33;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE Small conserved protein, ortholog of YAAA B-subtilis.
DCA0003.
GN CAC0003.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OC NCBI_TaxID=1488;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RC MEDLINE=21359325; PubMed=1146286;
RX Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,

```



RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing  
RT bacterium *Clostridium acetobutylicum*,"  
RL J. Bacteriol. 183:4823-4838(2001).  
DR EMBL; AE007513; AAK77990.1; -.  
DR F.R.; C96900.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR InterPro; IPR002942; S4.  
DR PROSITE; PS50889; S4; 1.  
KW Complete proteome.  
SQ SEQUENCE 68 AA; 7699 MW; 6AE9BB3FAA25753B CRC64;  
  
Query Match 100.0%; Score 25; DB 16; Length 68;  
Best Local Similarity 50.0%; Pred. No. 6.5e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 XEFIXDX 8  
Db :|||::|:  
8 TEFIKLDS 15  
  
RESULT 13  
Q8XPG0 PRELIMINARY; PRT; 68 AA.  
AC Q8XPG0;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein CPE0003.  
GN CPE0003.  
OS *Clostridium perfringens*.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1502;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=13 / Type A;  
RX MEDLINE=21664373; PubMed=11792842;  
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yanashita A.,  
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
RT "Complete genome sequence of *Clostridium perfringens*, an anaerobic  
RT flesh-eater,"  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
DR EMBL; AP003185; BAB79709.1; -.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR InterPro; IPR002942; S4.  
DR Pfam; PF01479; S4; 1.  
DR SMART; SM00363; S4; 1.  
DR PROSITE; PS50889; S4; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 68 AA; 7716 MW; 552AC70BCBC3D094 CRC64;  
  
Query Match 100.0%; Score 25; DB 16; Length 68;  
Best Local Similarity 50.0%; Pred. No. 6.5e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 XEFIXDX 8  
Db :|||::|:  
8 TEFIKLDS 15  
  
RESULT 14  
Q9F572 PRELIMINARY; PRT; 69 AA.  
AC Q9F572;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE YafA protein.  
GN YAF.  
OS *Escherichia coli*.

OG Plasmid R721.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K-12; TRANSPOSON=Tn7;  
RA Sampei G., Motomura K., Masuda S., Yamaguchi T., Ando K., Oishi T.,  
RA Furuya N., Komano T., Mizobuchi K.;  
RT "Organization and diversification of plasmid genomes: complete  
RT nucleotide sequence of the R721 genome,"  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K-12; TRANSPOSON=Tn7;  
RX MEDLINE=93015772; PubMed=1400257;  
RA Kim S., Komano T.;  
RT "Nucleotide sequence of the R721 shuffleon,"  
RL J. Bacteriol. 174:7053-7058(1992).  
DR EMBL; AP002527; BAB12596.1; -.  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
KW Plasmid.  
SQ SEQUENCE 69 AA; 7727 MW; E87C227EDD4B69D4 CRC64;  
  
Query Match 100.0%; Score 25; DB 2; Length 69;  
Best Local Similarity 50.0%; Pred. No. 6.6e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 XEFIXDX 8  
Db :|||::|:  
57 DEFIENDF 64  
  
RESULT 15  
Q8RHA8 PRELIMINARY; PRT; 71 AA.  
AC Q8RHA8;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical cytosolic protein FN2129.  
GN FN2129.  
OS *Fusobacterium nucleatum* (subsp. *nucleatum*).  
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;  
OC *Fusobacterium*.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;  
RX MEDLINE=21886394; PubMed=11889109;  
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,  
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
RA Fongstein M., Kyripides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium *Fusobacterium*  
RT *nucleatum* strain ATCC 25586,"  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL; AE010515; AAL94213.1; -.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR InterPro; IPR002942; S4.  
DR Pfam; PF01479; S4; 1.  
DR PROSITE; PS50889; S4; 1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 71 AA; 8334 MW; B255154730A57AB CRC64;  
  
Query Match 100.0%; Score 25; DB 16; Length 71;  
Best Local Similarity 50.0%; Pred. No. 6.8e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 XEFIXDX 8  
Db :|||::|:  
11 TEFIKLQ 18

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF410153; AAL69829.1; -;  
DR InterPro: IPR007952; Pox\_A3L.  
DR Pfam: PF05288; Pox\_A3L; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 75 AA; 8974 MW; EFF0DD892D47E3A CRC64;

Query Match 100.0%; Score 25; DB 12; Length 75;  
Best Local Similarity 50.0%; Pred. No. 7.2e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 25 YEFITDE 32

RESULT 18

Q91MS0 PRELIMINARY; PRT; 75 AA.

AC Q91MS0;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE LSDV093 hypothetical protein.  
GN LSDV093 OR LD093.  
OS Lumpy skin disease virus (LSDV).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Capripoxvirus.  
OX NCBI\_TaxID=59509;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Neethling 2490;  
RX MEDLINE=21329495; PubMed=11435593;  
RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Kutish G.F., Rock D.L.;  
RT "Genome of lumpy skin disease virus."  
RL J. Virol. 75:7122-7130(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Neethling 2490;  
RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Kutish G.F., Rock D.L.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Neethling Warmbaths LW;  
RA Kara P.D., Afonso C.L., Wallace D.B., Kutish G.F., Stipinovich C.,  
RA Lu Z., Vreede F.T., Taljaard L.C.F., Zsak A., Viljoen G.J., Rock D.L.;  
RT "Molecular characterization of the South African vaccine strain and  
RT the field isolate of lumpy skin disease virus."  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF325528; AAK85054.1; -;  
DR EMBL: AF409137; AAN02661.1; -;  
DR InterPro: IPR007952; Pox\_A3L.  
DR Pfam: PF05288; Pox\_A3L; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 75 AA; 9023 MW; 75C41C4CAF4FECFE CRC64;

Query Match 100.0%; Score 25; DB 12; Length 75;  
Best Local Similarity 50.0%; Pred. No. 7.2e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 25 YEFITDE 32

RESULT 19

Q8JTS8 PRELIMINARY; PRT; 75 AA.

ID Q8JTS8  
AC Q8JTS8;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein.

RESULT 16

Q31516 PRELIMINARY; PRT; 74 AA.

AC Q31516;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Mhc, class IIB protein (Fragment).  
GN MHC, CLASS IIB.  
OS Poecilia reticulata (Guppy).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
OC Cyprinodontiformes; Poeciliidae; Poecilia.  
OX NCBI\_TaxID=8081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96128251; PubMed=8537120;  
RA Sato A., Figueroa F., O'Huigin C., Reznick, Klein J.;  
RT "Identification of major histocompatibility complex genes in the  
RT guppy, Poecilia reticulata."  
RL Immunogenetics 43:38-49(1996).  
DR EMBL: Z54079; CAA90785.1; -;  
DR GO: GO:0016021; C: integral to membrane; IEA.  
DR GO: GO:0045012; F: MHC class II receptor activity; IEA.  
DR GO: GO:0019884; P: antigen presentation, exogenous antigen; IEA.  
DR GO: GO:0019886; P: antigen processing, exogenous antigen via M. .; IEA.  
DR GO: GO:0006955; P: immune response; IEA.  
DR InterPro: IPR000353; MHC II beta.  
DR Pfam: PF00969; MHC II beta; 1.  
DR ProDom: PD000328; MHC II beta; 1.  
KW Glycoprotein; MHC II; Transmembrane.  
FT NON\_TER 1 1  
FT NON\_TER 74 74  
SQ SEQUENCE 74 AA; 8585 MW; DD56AE2A8351987C CRC64;

Query Match 100.0%; Score 25; DB 7; Length 74;  
Best Local Similarity 50.0%; Pred. No. 7.1e+02;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|  
Db 24 LEFIRFDS 31

RESULT 17

Q8V3K6 PRELIMINARY; PRT; 75 AA.

AC Q8V3K6;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE SPV090 hypothetical protein.  
GN SPV090.  
OS Swinepox virus.  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Suipoxvirus.  
OX NCBI\_TaxID=10276;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=17077-99;  
RX MEDLINE=21624277; PubMed=11752168;  
RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Osorio F.A., Balinsky C.,  
RA Kutish G.F., Rock D.L.;  
RT "The genome of swinepox virus."  
RL J. Virol. 76:783-790(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=17077-99;  
RA Afonso C.L., Tulman E.R., Lu Z., Balinsky C., Osorio F.A., Zsak L.,  
RA Kutish G.F., Rock D.L.;

GN LW093.  
 OS Lumpy skin disease virus (LSDV).  
 CC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 CC Capripoxvirus.  
 OK NCBI\_TaxID=59509;  
 RN [1]\_TaxID=59509;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Neethling vaccine LW 1959;  
 RA Kara P.D., Afonso C.L., Wallace D.B., Kutish G.F., Stipinovich C.,  
 RA Lu Z., Vreede F.T., Taljaard L.C.F., Zsak A., Viljoen G.J., Rock D.L.;  
 RT "Molecular characterization of the South African vaccine strain and  
 RT the field isolate of lumpy skin disease virus.";  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF409138; AAN02818.1; -;  
 DR InterPro; IPR007952; PoxA3L;  
 DR Pfam; PF05286; PoxA3L; I.  
 KW Hypothetical protein.  
 SQ SEQUENCE 75 AA; 9022 MW; 98A41C4CAF4FECF6 CRC64;

Query Match 100.0%; Score 25; DB 12; Length 75;  
 Best Local Similarity 50.0%; Pred. No. 7.2e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 Db 25 FEFTEDE 32  
 :|||::|:

RESULT 20  
 Q84Q13 PRELIMINARY; PRT; 78 AA.  
 ID Q84Q13  
 AC Q84Q13  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DE REX1.S.  
 GN REX1.  
 OS Chlamydomonas reinhardtii.  
 CC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 CC Chlamydomonadales; Chlamydomonas.  
 OK NCBI\_TaxID=3055;  
 RN [1]\_TaxID=3055;  
 RP SEQUENCE FROM N.A.  
 RA Cenki B., Petersen J.L., Small G.D.;  
 RT "REX1, a novel gene required for DNA repair.";  
 RL J. Biol. Chem. 0:0-0(2003).  
 DR EMBL; AV236481; AAP1220.1; -;  
 SQ SEQUENCE 78 AA; 8941 MW; 34D47D31F8161F5D CRC64;

Query Match 100.0%; Score 25; DB 10; Length 78;  
 Best Local Similarity 50.0%; Pred. No. 7.5e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 Db 33 XEFITADL 40  
 :|||::|:

RESULT 21  
 Q97EC0 PRELIMINARY; PRT; 79 AA.  
 ID Q97EC0  
 AC Q97EC0;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein CAC3193.  
 GN CAC3193.  
 OS Clostridium acetobutylicum.  
 CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 CC Clostridium.  
 OK NCBI\_TaxID=1488;  
 RN [1]\_TaxID=1488;  
 RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
 EX MEDLINE=21359325; PubMed=11466286;  
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 RT bacterium Clostridium acetobutylicum.";  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL; AE007814; AAX81130.1; -;  
 DR PIR; G97292; G97292.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 79 AA; 9483 MW; 20426177BEEA12F5 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 79;  
 Best Local Similarity 50.0%; Pred. No. 7.6e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 Db 24 MEFINKDD 31  
 :|||::|:

RESULT 22  
 Q9FWI7 PRELIMINARY; PRT; 80 AA.  
 ID Q9FWI7  
 AC Q9FWI7;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN OSJNBA0065H03.5.  
 OS Oryza sativa (Rice).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 CC Ehrhartoideae; Oryzae; Oryza.  
 OK NCBI\_TaxID=4530;  
 RN [1]\_TaxID=4530;  
 RP SEQUENCE FROM N.A.  
 RA Wing R.A., Frisch D., Presting G., Wood T., Yu Y., Rambo T., Crane L.,  
 RA Schwartzbeck J., Thurmond S., Mao L.;  
 RT "Rice Genomic Sequence."  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC037197; AAG12482.2; -;  
 DR Gramene; Q9FWI7; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 80 AA; 8891 MW; 98F6D664F8B95709 CRC64;

Query Match 100.0%; Score 25; DB 10; Length 80;  
 Best Local Similarity 50.0%; Pred. No. 7.7e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 Db 32 VEFIAADM 39  
 :|||::|:

RESULT 23  
 Q7XHE3 PRELIMINARY; PRT; 80 AA.  
 ID Q7XHE3  
 AC Q7XHE3;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN OSJNBA0065H03.5.  
 OS Oryza sativa (japonica cultivar-group).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 CC Ehrhartoideae; Oryzae; Oryza.  
 OK NCBI\_TaxID=39947;  
 RN [1]\_TaxID=39947;  
 RP SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare;  
 RA The Rice Chromosome 10 Sequencing Consortium;  
 RT "In-depth view of structure, activity, and evolution of rice  
 chromosome 10.";  
 RL Science 300:1566-1569(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 DE EMBL; AE017048; AAP51792.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 80 AA; 8891 MW; 98F6D664F8B95709 CRC64;

Query Match 100.0%; Score 25; DB 10; Length 80;  
 Best Local Similarity 50.0%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||::|:  
 Db 32 VEFIADNM 39

## RESULT 24

Q88WY2 PRELIMINARY; PRT; 81 AA.  
 AC Q88WY2;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Molybdopter in biosynthesis protein, D chain.  
 GN MOAD OR IP 1479.  
 OS Lactobacillus plantarum.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 CC Lactobacillus.  
 OX NCBI\_TaxID=1590;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIMB 8826 / WCFS1;  
 RX MEDLINE=22480296; PubMed=12566566;  
 RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,  
 RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,  
 RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,  
 RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,  
 RA De Vos W.M., Sierzen R.J.;  
 RL "Complete genome sequence of Lactobacillus plantarum WCFS1.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).  
 DR EMBL; AL935256; CAD63935.1; -.  
 DR GO; GO:0006790; P:sulfur metabolism; IEA.  
 DR InterPro; IPR003749; This.  
 DR Pfam; PF02597; This; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 81 AA; 8545 MW; 35FB6D6AA1B97948 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 81;  
 Best Local Similarity 50.0%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||::|:  
 Db 56 QEFIADNR 63

## RESULT 25

Q81W38 PRELIMINARY; PRT; 81 AA.  
 AC Q81W38;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN BA4145.  
 OS Bacillus anthracis (strain Ames).

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=198094;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22608414; PubMed=12721629;  
 RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,  
 RA Nelson X.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
 RA Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,  
 RA Kolonay J.F., Beaman M.J., Dodson R.J., Brinkac L.M., Gwinn M.,  
 RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
 RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
 RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
 RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,  
 RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,  
 RA Fraser C.M.;  
 RT "The genome sequence of Bacillus anthracis Ames and comparison to  
 RT closely related bacteria.";  
 RL Nature 423:81-86(2003).  
 DR EMBL; AE017036; AAP27869.1; -.  
 DR TIGR; BA4145; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 81 AA; 9882 MW; 0933BB022A46379C CRC64;

Query Match 100.0%; Score 25; DB 16; Length 81;  
 Best Local Similarity 50.0%; Pred. No. 7.8e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||::|:  
 Db 6 MEFIKADE 13

## RESULT 26

Q812V8 PRELIMINARY; PRT; 82 AA.  
 AC Q812V8;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical cytosolic protein.  
 GN BC3935.  
 OS Bacillus cereus (strain ATCC 14579 / DSM 31).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=226900;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22608415; PubMed=12721630;  
 RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,  
 RA Kapatral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,  
 RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,  
 RA Grechkin Y., Pusch G., Haselkorn R., Fongstein M., Ehrlich S.D.,  
 RA Overbeek R., Kyrpides N.;  
 RT "Genome sequence of Bacillus cereus and comparative analysis with  
 RT Bacillus anthracis.";  
 RL Nature 423:87-91(2003).  
 DR EMBL; AE017010; AAP10855.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 82 AA; 10000 MW; 1C1104962489B6AC CRC64;

Query Match 100.0%; Score 25; DB 16; Length 82;  
 Best Local Similarity 50.0%; Pred. No. 7.9e+02;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
 :|||::|:  
 Db 7 MEFIKADE 14

## RESULT 27

Q95HS2 PRELIMINARY; PRT; 85 AA.  
 ID Q95HS2  
 AC Q95HS2;

```

DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE MHC class II antigen beta chain (Fragment).
GN PEMA-EB.
OS Peromyscus maniculatus (Deer mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Sigmodontinae;
OC Peromyscus.
OX NCBI_TaxID=10042;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SQPM26a;
RA Richman A.D., Herrera L.G., Nash D.;
RT "MHC Class II Beta Sequence Diversity in the Deermouse (Peromyscus
RL maniculatus): Implications for Models of Balancing Selection.";
Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF312753; AAK98053.1; -.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0045012; F: MHC class II receptor activity; IEA.
DR GO: GO:0019884; P: antigen presentation, exogenous antigen; IEA.
DR GO: GO:0019886; P: antigen processing, exogenous antigen via M. . .; IEA.
DR GO: GO:0006955; P: immune response; IEA.
DR InterPro: IPR000353; MHC_II_beta.
DR Pfam: PF00969; MHC_II_beta; 1.
DR ProDom: PD000328; MHC_II_beta; 1.
KW MHC.
FT NON_TER 1 1
FT NON_TER 85 85
SQ SEQUENCE 85 AA; 10237 MW; 3061A008A7BE71DB CRC64;

Query Match 100.0%; Score 25; DB 7; Length 85;
Best Local Similarity 50.0%; Pred. No. 8.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 26 EEFIFDS 33

RESULT 28
Q8UIA6 PRELIMINARY; PRT; 85 AA.
ID Q8UIA6
AC Q8UIA6;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein PF1315.
GN PF1315.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE010237; AAL81439.1; -.
DR InterPro: IPR004919; DUF262.
DR Pfam: PF03235; DUF262; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 85 AA; 9560 MW; B3CE50A6CFF33C4F CRC64;

Query Match 100.0%; Score 25; DB 17; Length 85;
Best Local Similarity 50.0%; Pred. No. 8.2e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 56 EEFITDDP 63

RESULT 29
O69179 PRELIMINARY; PRT; 86 AA.
ID O69179;
AC O69179;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Putative hydroxylase component.
GN PHLL.
OS Alcaligenes eutrophus (Ralstonia eutropha).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JMP134;
RX PubMed=9797289;
RA Ayoubi P.J., Harker A.R.;
RT "Whole-cell kinetics of trichloroethylene degradation by phenol
RT hydroxylase in a Ralstonia eutropha JMP134 derivative.";
Appl. Environ. Microbiol. 64:4353-4356 (1998).
DR EMBL: AF065891; AAC77381.1; -.
SQ SEQUENCE 86 AA; 9583 MW; 6F9E4D89D5C00840 CRC64;

Query Match 100.0%; Score 25; DB 2; Length 86;
Best Local Similarity 50.0%; Pred. No. 8.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 79 LEFIFDA 86

RESULT 30
Q51940 PRELIMINARY; PRT; 86 AA.
ID Q51940
AC Q51940;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein (Tbc2B monooxygenase).
GN TBUU OR TBC2B.
OS Burkholderia pickettii (Pseudomonas pickettii), and
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=329, 292;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=B.pickettii; STRAIN=PK01;
RX MEDLINE=95172404; PubMed=7867951;
RA Byrne A.M., Kukor J.J., Olsen R.H.;
RT "Sequence analysis of the gene cluster encoding toluene-3-
RT monooxygenase from Pseudomonas pickettii PK01.";
Gene 134:65-70 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=B.cepacia; STRAIN=JS150;
RA Kahng H.Y., Kukor J.J.;
RT "Genetic and functional analysis of Tbc monooxygenases for catabolism
RT of alkyl- and chloroaromatic compounds in Burkholderia cepacia
RT JS150.";
Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: U04052; AAB09619.1; -.
DR EMBL: AF282898; AAG40795.1; -.
DR GO: GO:0004497; F: monooxygenase activity; IEA.
KW Hypothetical protein; Monooxygenase.
SQ SEQUENCE 86 AA; 9604 MW; 7D50901A65635AB3 CRC64;

Query Match 100.0%; Score 25; DB 2; Length 86;
Best Local Similarity 50.0%; Pred. No. 8.3e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```
QY 1 XEFIXDX 8
   :|||:|:|:
Db 79 LEFICDA 86

RESULT 31
Q8PUJ6 Q8PUJ6 PRELIMINARY; PRT; 90 AA.
AC Q8PUJ6;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein XAC2631.
GN XAC2631.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
[1]
SEQUENCE FROM N.A.
STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2202145; PubMed=12024217;
da Silva A.C.R., Ferro J.A., Reinach F.C., Fatah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Canarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Teai S.M., White F.F.,
RA Setubal J.C., Kikajima J.P.;
RA "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463(2002).
DR EMBL: AE011903; AAM37478.1; -.
RX Hypothetical protein: Complete proteome.
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 90 AA; 10209 MW; D4F880AED6935695 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 90;
Best Local Similarity 50.0%; Pred. NO. 8.7e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
   :|||:|:|:
Db 29 QEFIELD 36

RESULT 32
Q8ET09 Q8ET09 PRELIMINARY; PRT; 90 AA.
AC Q8ET09;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical conserved protein.
GN OB0456.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
[1]
SEQUENCE FROM N.A.
STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme

environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL: AP004594; BAC12412.1; -.
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 90 AA; 10509 MW; D2647123061E723B CRC64;

Query Match 100.0%; Score 25; DB 16; Length 90;
Best Local Similarity 50.0%; Pred. NO. 8.7e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
   :|||:|:|:
Db 8 HEPIARDV 15

RESULT 33
Q9GJY0 Q9GJY0 PRELIMINARY; PRT; 91 AA.
AC Q9GJY0;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Fructose-1,6-bisphosphatase (EC 3.1.3.11) (Fragment).
GN FBPI.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
[1]
SEQUENCE FROM N.A.
RA Tillmann H., Eschrich K.;
RT "Liver and muscle fructose-1,6-bisphosphatase isoenzyme amounts in
RT different rabbit tissues.";
RL Submitted (DRC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ300657; CAC19334.1; -.
RX HSP; P00637; IBA4.
DR GO: GO:0042132; F:fructose-bisphosphatase activity; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0042578; F:phosphoric ester hydrolase activity; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro: IPR000146; In_FB_phptase.
DR Pfam: PF00316; FBPase; 1.
DR ProDom: PD001491; In_FB_phptase; 1.
KW Hydrolase.
FT NON_TER 1 91
FT NON_TER 91 91
SQ SEQUENCE 91 AA; 9919 MW; 3D9C4C087C09AA33 CRC64;

Query Match 100.0%; Score 25; DB 6; Length 91;
Best Local Similarity 50.0%; Pred. NO. 8.8e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
   :|||:|:|:
Db 39 GEFLVADK 46

RESULT 34
Q7TIF7 Q7TIF7 PRELIMINARY; PRT; 92 AA.
AC Q7TIF7;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
[1]
SEQUENCE FROM N.A.
RA Ouyang J., Nie Q., Zhang X.;
```

```

RT "Sequence of cGHR intron 9."
RL EMBL; AY327492; AAP92122.1; -.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 92 92
SQ SEQUENCE 92 AA; 10377 MW; 5F624696942A3269 CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 13; Length 92;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 48 VEFIELDI 55

RESULT 35
Q93EH4
ID Q93EH4 PRELIMINARY; PRT; 95 AA.
AC Q93EH4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Flagellar biosynthesis protein (fragment).
GN FLHA.
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3B1;
RA Ge Z.; Feng Y.; Fox J.G.;
RT "Helicobacter hepaticus genome: construction of an ordered cosmid
library and sequence analysis of the selected genomic regions."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF358691; AAL16680.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0009306; P:protein secretion; IEA.
DR InterPro; IPR001712; Bact_FH1PEP.
DR Pfam; PF00771; FH1PEP; 1.
FT NON_TER 1 1
FT NON_TER 95 95
SQ SEQUENCE 95 AA; 10789 MW; BED4A41B153BEA53 CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 2; Length 95;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 10 EEFITXDE 17

RESULT 36
Q24060
ID Q24060 PRELIMINARY; PRT; 95 AA.
AC Q24060;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ecdysone-inducible protein 71EJ.
GN EIG71EJ OR L71-10 OR CG7588.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RA MEDLINE=96152797; PubMed=8568884;

RT Wright L.G., Chen T., Thummel C.S., Guild G.M.;
RT "Molecular characterization of the 71E late puff in Drosophila
melanogaster reveals a family of novel genes."
RL J. Mol. Biol. 255:387-400(1996).
DR EMBL; U23836; AAA74183.1; -.
DR FlyBase; FBgn0014850; Big71EJ.
DR InterPro; IPR003475; Insect_Unk.
DR Pfam; PF02448; L71; 1.
SQ SEQUENCE 95 AA; 11446 MW; 33171F38EA2826D1 CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 5; Length 95;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 46 LEFINSDC 53

RESULT 37
Q814L7
ID Q814L7 PRELIMINARY; PRT; 95 AA.
AC Q814L7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein E02A10.4.
GN E02A10.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Thomas K.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z81053; CAD54130.1; -.
DR WormPep; E02A10.4; CE31898.
KW Hypothetical protein.
SQ SEQUENCE 95 AA; 10397 MW; 46723F3FEBBF6189 CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 5; Length 95;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 27 EEFIAVDC 34

RESULT 38
Q8D9P7
ID Q8D9P7 PRELIMINARY; PRT; 95 AA.
AC Q8D9P7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Plasmid stabilization system protein.
GN WV12547.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCF6;

```

RA Adams M.D., Cerniker S.E., Holt R.A., Evans C.R., Gocayne J.D.  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.  
RA Adams M.D., Cerniker S.E., Holt R.A., Evans C.R., Gocayne J.D.

RA Adams M.D., Cerniker S.E., Holt R.A., Evans C.R., Gocayne J.D.  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.  
RA Adams M.D., Cerniker S.E., Holt R.A., Evans C.R., Gocayne J.D.



```

RX MEDLINE=22617848; PubMed=12732969;
RA Xu Y., Chen M., Zhang W., Lin M.;
RT "Genetic organization of genes encoding phenol hydroxylase, benzoate
RT 1,2-dioxygenase alpha subunit and its regulatory proteins in
RT Acinetobacter calcoaceticus PHEA-2.";
RL Curr. Microbiol. 46:235-240(2003).
DR EMBL: AJ564846; CAD92311.1; -.
SQ SEQUENCE 99 AA; 11542 MW; 11BF76425EA2D847 CRC64;

Query Match          100.0%; Score 25; DB 2; Length 99;
Best Local Similarity 50.0%; Pred. No. 9.7e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 25 AEFIEFDF 32

RESULT 42
Q863N9 PRELIMINARY; PRT; 101 AA.
AC Q863N9;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
GN GHR.
OS Arctonyx collaris.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Melinae;
OC Arctonyx.
OX NCBI_TaxID=139309;
RN [1]
RP SEQUENCE FROM N.A.
RA Koepfli K.-P., Wayne R.K.;
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";
RL Syst. Biol. 0:0-0(2003).
DR EMBL: AF498204; AAP19695.1; -.
DR GO: GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 101 101
SQ SEQUENCE 101 AA; 11428 MW; AD696CB6E08BD7AC CRC64;

Query Match          100.0%; Score 25; DB 6; Length 101;
Best Local Similarity 50.0%; Pred. No. 9.9e+02;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 28 VEFIELDI 35

RESULT 43
Q92528 PRELIMINARY; PRT; 102 AA.
AC Q92528;
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE 11.6 kDa protein.
OS Carnation latent virus (CLV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_TaxID=12164;
RN [1]
RP SEQUENCE FROM N.A.
RA Meehan B.M.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91324119; PubMed=1713905;
RA Meehan B.M., Mills P.R.;

```

```

RT "Nucleotide sequence of the 3'-terminal region of carnation latent
RT virus.";
RL Intervirology 32:262-267(1991).
DR EMBL: AJ010697; CAA09307.1; -.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR002568; Carla_C4.
DR Pfam: PF01623; Carla_C4; 1.
SQ SEQUENCE 102 AA; 11584 MW; 4574B3FF43B46B19 CRC64;

Query Match          100.0%; Score 25; DB 12; Length 102;
Best Local Similarity 50.0%; Pred. No. 1e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 90 LEFIGIDL 97

RESULT 44
Q8UJL1 PRELIMINARY; PRT; 102 AA.
AC Q8UJL1;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative nucleic acid binding protein.
OS Carnation latent virus (CLV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_TaxID=12164;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91324119; PubMed=1713905;
RA Meehan B.M., Mills P.R.;
RT "Nucleotide sequence of the 3'-terminal region of carnation latent
RT virus.";
RL Intervirology 32:262-267(1991).
DR EMBL: X52627; CAA36855.1; -.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR002568; Carla_C4.
DR Pfam: PF01623; Carla_C4; 1.
SQ SEQUENCE 102 AA; 11626 MW; 0474AF8F5FD81B05 CRC64;

Query Match          100.0%; Score 25; DB 12; Length 102;
Best Local Similarity 50.0%; Pred. No. 1e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
Db 90 LEFIGIDL 97

RESULT 45
Q88LJ6 PRELIMINARY; PRT; 102 AA.
AC Q88LJ6;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Transcriptional regulator, Cro/CI family.
GN P1935.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,

```

```

RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";
RL Syst. Biol. 0:0-0(2003).
DR EMBL: AF498187; AAP19678.1; -.
DR GO: GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON TER 1
FT NON TER 105
SQ SEQUENCE 105 AA; 11852 MW; F07A7052451EB8E7 CRC64;

Query Match 100.0%; Score 25; DB 6; Length 105;
Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;
Matches 4; Conservative 4; Mismatches 4;

QY 1 XEFIXDX 8
Db 32 VEFIELDI 39

RESULT 48
Q863P4 PRELIMINARY; PRT; 105 AA.
ID Q863P4;
AC Q863P4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
GN GHR.
OS Eira barbara (Tayra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;
OC Eira.
OX NCBI_TaxID=204263;
RN [1]_TaxID=204263;
RP SEQUENCE FROM N.A.
RX MEDLINE=20261044; PubMed=1083899;
RA Oshida K., Kuriki T., Takata H., Okada S.;
RT "Cloning of the cyclodextrin glucanotransferase gene from alkalophilic
RL Bacillus sp. A2-5a and analysis of the raw starch-binding domain.";
RT Appl. Microbiol. Biotechnol. 53:430-434(2000).
DR EMBL: AB015670; BAA31532.1; -.
DR GO: GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON TER 1
FT NON TER 105
SQ SEQUENCE 105 AA; 11879 MW; F6F6445250DB79BF CRC64;

Query Match 100.0%; Score 25; DB 6; Length 105;
Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;
Matches 4; Conservative 4; Mismatches 4;

QY 1 XEFIXDX 8
Db 32 VEFIELDI 39

RESULT 49
Q8C3V0 PRELIMINARY; PRT; 105 AA.
ID Q8C3V0;
AC Q8C3V0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Lung;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of

Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA Uterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808(2002).
DR EMBL: AE016781; AAN67552.1; -.
DR TIGR: PP1935; -.
DR GO: GO:0003677; F:DNA binding; IEA.
DR InterPro: IPR001387; HTH_3.
DR Pfam: PF01381; HTH_3; 1.
KW Complete proteome.
SQ SEQUENCE 102 AA; 11480 MW; 9A2E2D606938A8A5 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 102;
Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;
Matches 4; Conservative 4; Mismatches 4;

QY 1 XEFIXDX 8
Db 78 AEFINGDP 85

RESULT 46
Q82973 PRELIMINARY; PRT; 105 AA.
ID Q82973;
AC Q82973;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE A2-5a orf4.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]_TaxID=1409;
RP SEQUENCE FROM N.A.
RX MEDLINE=20261044; PubMed=1083899;
RA Oshida K., Kuriki T., Takata H., Okada S.;
RT "Cloning of the cyclodextrin glucanotransferase gene from alkalophilic
RL Bacillus sp. A2-5a and analysis of the raw starch-binding domain.";
RT Appl. Microbiol. Biotechnol. 53:430-434(2000).
DR EMBL: AB015670; BAA31532.1; -.
DR GO: GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON TER 1
FT NON TER 105
SQ SEQUENCE 105 AA; 12243 MW; 9677A516E4C4916B CRC64;

Query Match 100.0%; Score 25; DB 2; Length 105;
Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;
Matches 4; Conservative 4; Mismatches 4;

QY 1 XEFIXDX 8
Db 82 NEFIGFDI 89

RESULT 47
Q863Q6 PRELIMINARY; PRT; 105 AA.
ID Q863Q6;
AC Q863Q6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
GN GHR.
OS Lontra canadensis (River otter) (Lutra canadensis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Lutrinae;
OC Lontra.
OX NCBI_TaxID=76717;
RN [1]_TaxID=76717;
RP SEQUENCE FROM N.A.
RX Koepfli K.-P., Wayne R.K.;
RA "Type-1 SIS Markers Are More Informative Than Cytochrome b in

```

```
RT 60,770 full-length cDNAs." ;
RL Nature 420:563-573 (2002).
DR EMBL; AK084868; BAC39298.1; -.
KW Hypothetical protein.
SQ SEQUENCE 105 AA; 12193 MW; B1CBD28D2FB01797 CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 11; Length 105;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
   :|||:
Db 79 LEFINEDY 86

RESULT 50
Q863Q3
ID Q863Q3 PRELIMINARY; PRT; 106 AA.
AC Q863Q3;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
GN GHR.
OS Lutra lutra (European river otter).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Lutrinae;
OC Lutra.
OX NCBI_TaxID=9657;
RN [1]
RP SEQUENCE FROM N.A.
RA Koepfli K.-P., Wayne R.K.;
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";
RL Syst. Biol. 0:0-0(2003).
DR EMBL; AF498190; AAP19681.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON TER 1
FT NON TER 106
SQ SEQUENCE 106 AA; 12006 MW; EFB0C6D8793AE9FF CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 6; Length 106;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
   :|||:
Db 33 VEFIELDI 40

RESULT 51
Q863Q2
ID Q863Q2 PRELIMINARY; PRT; 106 AA.
AC Q863Q2;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
GN GHR.
OS Lutra maculicollis (Spotted necked otter).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Lutrinae;
OC Lutra.
OX NCBI_TaxID=76719;
RN [1]
RP SEQUENCE FROM N.A.
RA Koepfli K.-P., Wayne R.K.;
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";
RL Syst. Biol. 0:0-0(2003).
DR EMBL; AF498191; AAP19682.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
```

```
KW Receptor.
FT NON TER 1
FT NON TER 106
SQ SEQUENCE 106 AA; 12003 MW; EFB0DDAEA24C29FF CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 6; Length 106;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
   :|||:
Db 33 VEFIELDI 40

RESULT 52
Q863Q9
ID Q863Q9 PRELIMINARY; PRT; 107 AA.
AC Q863Q9;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
GN GHR.
OS Anonyx capensis (Cape clawless otter).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Lutrinae;
OC Anonyx.
OX NCBI_TaxID=76722;
RN [1]
RP SEQUENCE FROM N.A.
RA Koepfli K.-P., Wayne R.K.;
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";
RL Syst. Biol. 0:0-0(2003).
DR EMBL; AF498184; AAP19675.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON TER 1
FT NON TER 107
SQ SEQUENCE 107 AA; 12101 MW; 7F3D579D3C26011B CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 6; Length 107;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
   :|||:
Db 34 VEFIELDI 41

RESULT 53
Q863Q8
ID Q863Q8 PRELIMINARY; PRT; 107 AA.
AC Q863Q8;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
GN GHR.
OS Amblyonyx cinereus (Asian small-clawed otter) (Anonyx cinereus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Lutrinae;
OC Amblyonyx.
OX NCBI_TaxID=55043;
RN [1]
RP SEQUENCE FROM N.A.
RA Koepfli K.-P., Wayne R.K.;
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";
RL Syst. Biol. 0:0-0(2003).
DR EMBL; AF498185; AAP19676.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
```

```

FT NON_TER 107 107
SQ SEQUENCE 107 AA; 12180 MW; 9113E140F3C5C3BB CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 6; Length 107;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 34 VEFIELDI 41

RESULT 56
Q863Q4 PRELIMINARY; PRT; 107 AA.
AC Q863Q4; 2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
GN GHR.
OS Lontra longicaudis (Neotropical otter) (Lutra longicaudis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Lutrinae;
OC Lontra.
OX NCBI_TaxID=71113;
RN [1]
RP SEQUENCE FROM N.A.
RA Koepfli K.-P., Wayne R.K.;
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";
RL Syst. Biol. 0:0-0(2003).
DR EMBL; AF498189; AAP19680.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 107 107
SQ SEQUENCE 107 AA; 12166 MW; 9113E146451EA8BB CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 6; Length 107;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 34 VEFIELDI 41

RESULT 57
Q863Q1 PRELIMINARY; PRT; 107 AA.
AC Q863Q1; 2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
GN GHR.
OS Pteronura brasiliensis (Giant otter).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Lutrinae;
OC Pteronura.
OX NCBI_TaxID=9672;
RN [1]
RP SEQUENCE FROM N.A.
RA Koepfli K.-P., Wayne R.K.;
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";
RL Syst. Biol. 0:0-0(2003).
DR EMBL; AF498192; AAP19683.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 107 107

```

```

FT NON_TER 1 1
FT NON_TER 107 107
SQ SEQUENCE 107 AA; 12121 MW; 84DD579D3C2EB8BA CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 6; Length 107;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 34 VEFIELDI 41

RESULT 54
Q863Q7 PRELIMINARY; PRT; 107 AA.
AC Q863Q7; 2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
GN GHR.
OS Enhydra lutris (Sea otter).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Lutrinae;
OC Enhydra.
OX NCBI_TaxID=34882;
RN [1]
RP SEQUENCE FROM N.A.
RA Koepfli K.-P., Wayne R.K.;
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";
RL Syst. Biol. 0:0-0(2003).
DR EMBL; AF498186; AAP19677.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 107 107
SQ SEQUENCE 107 AA; 12151 MW; 84DD579D3C2EA9FE CRC64;

Query Match
Best Local Similarity 100.0%; Score 25; DB 6; Length 107;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 34 VEFIELDI 41

RESULT 55
Q863Q5 PRELIMINARY; PRT; 107 AA.
AC Q863Q5; 2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Growth hormone receptor (Fragment).
GN GHR.
OS Lontra felina (Sea cat) (Lutra felina).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Lutrinae;
OC Lontra.
OX NCBI_TaxID=76718;
RN [1]
RP SEQUENCE FROM N.A.
RA Koepfli K.-P., Wayne R.K.;
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";
RL Syst. Biol. 0:0-0(2003).
DR EMBL; AF498188; AAP19679.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON_TER 1 1
FT NON_TER 107 107

```

SQ SEQUENCE 107 AA; 12137 MW; 96DCA309E7789FE CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;  
Best Local Similarity 50.0%; Pred. No. 1e+03; 0; Indels 0; Gaps 0;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db :||||:|  
34 VEFIELDI 41

RESULT 58

Q863Q0 ID Q863Q0 PRELIMINARY; PRT; 107 AA.  
AC Q863Q0;  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Growth hormone receptor (Fragment).  
GN GHR.  
OS Mustela erminea (Ermine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;  
OC Mustela.  
OX NCBI\_TaxID=36723;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koepfli K.-P., Wayne R.K.;  
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in  
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";  
RL Syst. Biol. 0:0-0(2003).  
DR EMBL; AF498193; AAP19684.1; -.  
DR GO; GO:0004872; F:receptor activity; IEA.  
KW Receptor.  
FT NON\_TER 1 1  
FT NON\_TER 107 107  
SQ SEQUENCE 107 AA; 12140 MW; 96DBD146450149FE CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;  
Best Local Similarity 50.0%; pred. No. 1e+03; 0; Indels 0; Gaps 0;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db :||||:|  
34 VEFIELDI 41

RESULT 59

Q863P9 ID Q863P9 PRELIMINARY; PRT; 107 AA.  
AC Q863P9;  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Growth hormone receptor (Fragment).  
GN GHR.  
OS Mustela frenata (Long-tailed weasel).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;  
OC Mustela.  
OX NCBI\_TaxID=55048;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koepfli K.-P., Wayne R.K.;  
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in  
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";  
RL Syst. Biol. 0:0-0(2003).  
DR EMBL; AF498194; AAP19685.1; -.  
DR GO; GO:0004872; F:receptor activity; IEA.  
KW Receptor.  
FT NON\_TER 1 1  
FT NON\_TER 107 107  
SQ SEQUENCE 107 AA; 12140 MW; 96DBD146450149FE CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;  
Best Local Similarity 50.0%; Pred. No. 1e+03; 0; Indels 0; Gaps 0;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db :||||:|  
34 VEFIELDI 41

RESULT 60

Q863P8 ID Q863P8 PRELIMINARY; PRT; 107 AA.  
AC Q863P8;  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Growth hormone receptor (Fragment).  
GN GHR.  
OS Mustela vison (American mink).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;  
OC Mustela.  
OX NCBI\_TaxID=9667;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koepfli K.-P., Wayne R.K.;  
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in  
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";  
RL Syst. Biol. 0:0-0(2003).  
DR EMBL; AF498195; AAP19686.1; -.  
DR GO; GO:0004872; F:receptor activity; IEA.  
KW Receptor.  
FT NON\_TER 1 1  
FT NON\_TER 107 107  
SQ SEQUENCE 107 AA; 12149 MW; 96CCE146450149FE CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;  
Best Local Similarity 50.0%; pred. No. 1e+03; 0; Indels 0; Gaps 0;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
Db :||||:|  
34 VEFIELDI 41

RESULT 61

Q863P7 ID Q863P7 PRELIMINARY; PRT; 107 AA.  
AC Q863P7;  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Growth hormone receptor (Fragment).  
GN GHR.  
OS Martes americana (American marten).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;  
OC Martes.  
OX NCBI\_TaxID=9660;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koepfli K.-P., Wayne R.K.;  
RT "Type-1 STS Markers Are More Informative Than Cytochrome b in  
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";  
RL Syst. Biol. 0:0-0(2003).  
DR EMBL; AF498196; AAP19687.1; -.  
DR GO; GO:0004872; F:receptor activity; IEA.  
KW Receptor.  
FT NON\_TER 1 1  
FT NON\_TER 107 107  
SQ SEQUENCE 107 AA; 12172 MW; 96DBD14658B649FE CRC64;

Mon Sep 13 10:53:34 2004

us-09-660-302e-1.rspt

Query Match 100.0%; Score 25; DB 6; Length 107;  
Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEPIXXDX 8  
:||||:|  
Db 34 VEFIELDI 41

## RESULT 62

Q863P6 PRELIMINARY; PRT; 107 AA.  
AC Q863P6; (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Growth hormone receptor (Fragment).  
GN GHR.  
OS Martes pennanti (Marten).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;  
OC Martes.  
OC NCBI\_TaxID=76720;  
RN [1] SEQUENCE FROM N.A.  
RP Koepfli K.-P., Wayne R.K.;  
RA "Type-1 STS Markers Are More Informative Than Cytochrome b in  
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";  
RL Syst. Biol. 0:0-0(2003).  
DR EMBL; AF498197; AAP19688.1; --  
DR GO; GO:0004872; F:receptor activity; IEA.  
KW Receptor.  
FT NON TER 1  
FT NON TER 107  
SQ SEQUENCE 107 AA; 12136 MW; 830606D83B0149FE CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;  
Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEPIXXDX 8  
:||||:|  
Db 34 VEFIELDI 41

## RESULT 63

Q863P5 PRELIMINARY; PRT; 107 AA.  
AC Q863P5; (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Growth hormone receptor (Fragment).  
GN GHR.  
OS Gulo gulo (Wolverine) (Gluton).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;  
OC Gulo.  
OC NCBI\_TaxID=48420;  
RN [1] SEQUENCE FROM N.A.  
RP Koepfli K.-P., Wayne R.K.;  
RA "Type-1 STS Markers Are More Informative Than Cytochrome b in  
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";  
RL Syst. Biol. 0:0-0(2003).  
DR EMBL; AF498198; AAP19689.1; --  
DR GO; GO:0004872; F:receptor activity; IEA.  
KW Receptor.  
FT NON TER 1  
FT NON TER 107  
SQ SEQUENCE 107 AA; 12140 MW; 96DBD146450149FE CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;  
Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;

QY 1 XEPIXXDX 8  
:||||:|  
Db 34 VEFIELDI 41

Query Match 100.0%; Score 25; DB 6; Length 107;  
Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEPIXXDX 8  
:||||:|  
Db 34 VEFIELDI 41

## RESULT 64

Q863P3 PRELIMINARY; PRT; 107 AA.  
AC Q863P3; (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Growth hormone receptor (Fragment).  
GN GHR.  
OS Galictis vittata.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;  
OC Galictis.  
OC NCBI\_TaxID=204265;  
RN [1] SEQUENCE FROM N.A.  
RP Koepfli K.-P., Wayne R.K.;  
RA "Type-1 STS Markers Are More Informative Than Cytochrome b in  
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";  
RL Syst. Biol. 0:0-0(2003).  
DR EMBL; AF498200; AAP19691.1; --  
DR GO; GO:0004872; F:receptor activity; IEA.  
KW Receptor.  
FT NON TER 1  
FT NON TER 107  
SQ SEQUENCE 107 AA; 12109 MW; 4048E64F61DBF52C CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;  
Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEPIXXDX 8  
:||||:|  
Db 34 VEFIELDI 41

## RESULT 65

Q863P2 PRELIMINARY; PRT; 107 AA.  
AC Q863P2; (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Growth hormone receptor (Fragment).  
GN GHR.  
OS Ictonyx striatus (striped polecat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;  
OC Ictonyx.  
OC NCBI\_TaxID=55050;  
RN [1] SEQUENCE FROM N.A.  
RP Koepfli K.-P., Wayne R.K.;  
RA "Type-1 STS Markers Are More Informative Than Cytochrome b in  
RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";  
RL Syst. Biol. 0:0-0(2003).  
DR EMBL; AF498201; AAP19692.1; --  
DR GO; GO:0004872; F:receptor activity; IEA.  
KW Receptor.  
FT NON TER 1  
FT NON TER 107  
SQ SEQUENCE 107 AA; 12140 MW; 96DBD146450149FE CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;  
Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;

QY 1 XEPIXXDX 8  
:||||:|  
Db 34 VEFIELDI 41

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:|||||:  
Db 34 VEFIELDI 41

RESULT 66

Q863P1 PRELIMINARY; PRT; 107 AA.

AC Q863P1; (Tremblrel. 24, Created)

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Growth hormone receptor (Fragment).

GN GHR.

OS Meles meles (Eurasian badger).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Melinae; Meles.

OX NCBI\_TaxID=9662;

RN [1]

RP SEQUENCE FROM N.A.

RA Koepfli K.-P., Wayne R.K.;

RT "Type-1 STS Markers Are More Informative Than Cytochrome b in

RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";

RL Syst. Biol. 0:0-0(2003).

DR EMBL; AF498202; AAP19693.1; --

DR GO; GO:0004872; F:receptor activity; IEA.

KW Receptor.

FT NON\_TER 1 1

FT NON\_TER 107 107

SQ SEQUENCE 107 AA; 12110 MW; 96DBD15D44B149FE CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;

Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:|||||:  
Db 34 VEFIELDI 41

RESULT 67

Q863P0 PRELIMINARY; PRT; 107 AA.

AC Q863P0; (Tremblrel. 24, Created)

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Growth hormone receptor (Fragment).

GN GHR.

OS Taxidea taxus (Badger).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Taxidiinae;

OX NCBI\_TaxID=30554;

RN [1]

RP SEQUENCE FROM N.A.

RA Koepfli K.-P., Wayne R.K.;

RT "Type-1 STS Markers Are More Informative Than Cytochrome b in

RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";

RL Syst. Biol. 0:0-0(2003).

DR EMBL; AF498203; AAP19694.1; --

DR GO; GO:0004872; F:receptor activity; IEA.

KW Receptor.

FT NON\_TER 1 1

FT NON\_TER 107 107

SQ SEQUENCE 107 AA; 12113 MW; 825D0A3F750149FE CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;

Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:|||||:  
Db 34 VEFIELDI 41

RESULT 68

Q863N8 PRELIMINARY; PRT; 107 AA.

AC Q863N8; (Tremblrel. 24, Created)

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Growth hormone receptor (Fragment).

GN GHR.

OS Melogale moschata.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Melinae;

OC Melogale.

OX NCBI\_TaxID=204267;

RN [1]

RP SEQUENCE FROM N.A.

RA Koepfli K.-P., Wayne R.K.;

RT "Type-1 STS Markers Are More Informative Than Cytochrome b in

RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";

RL Syst. Biol. 0:0-0(2003).

DR EMBL; AF498205; AAP19696.1; --

DR GO; GO:0004872; F:receptor activity; IEA.

KW Receptor.

FT NON\_TER 1 1

FT NON\_TER 107 107

SQ SEQUENCE 107 AA; 12141 MW; C671DBEC450C29F3 CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;

Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:|||||:  
Db 34 VEFIELDI 41

RESULT 69

Q863N7 PRELIMINARY; PRT; 107 AA.

AC Q863N7; (Tremblrel. 24, Created)

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Growth hormone receptor (Fragment).

GN GHR.

OC Bassariscus astutus (ringtail).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Procyonidae; Bassariscus.

OX NCBI\_TaxID=55047;

RN [1]

RP SEQUENCE FROM N.A.

RA Koepfli K.-P., Wayne R.K.;

RT "Type-1 STS Markers Are More Informative Than Cytochrome b in

RT Phylogenetic Reconstruction of the Mustelidae (Mammalia: Carnivora).";

RL Syst. Biol. 0:0-0(2003).

DR EMBL; AF498206; AAP19697.1; --

DR GO; GO:0004872; F:receptor activity; IEA.

KW Receptor.

FT NON\_TER 1 1

FT NON\_TER 107 107

SQ SEQUENCE 107 AA; 12055 MW; 977B2547E01459BA CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;

Best Local Similarity 50.0%; Pred. No. 1e+03; Indels 0; Gaps 0;

Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:|||||:  
Db 34 VEFIELDI 41

```

Db      34 VEFIELDI 41

RESULT 70
Q8XYS2 PRELIMINARY; PRT; 107 AA.
AC Q8XYS2;
ID Q863N6;
IT Q863N6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DE Probable phage HK022 GP9-related protein.
DN RSC1686 OR RS04080.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OX Burkholderiaceae; Ralstonia.
NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.-C., Cattolico L.,
RA Chandler M., Cholsne N., Claudel-Renaud C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Signier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646065; CAD15388.1; -.
DR InterPro; IPR006453; Gpi6_SPP1.
DR TIGRFAMs; TIGR01563; gpi6_SPP1; 1.
KW Complete proteome.
SQ SEQUENCE 107 AA; 12024 MW; 96CB24F6501459BB CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;
Best Local Similarity 50.0%; Pred. No. 1e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIYXXDX 8
   :|::||:
Db      34 VEFIELDI 41

RESULT 71
Q8DT70 PRELIMINARY; PRT; 107 AA.
AC Q8DT70;
ID Q863N6;
IT Q863N6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DE Probable phage HK022 GP9-related protein.
DN RSC1686 OR RS04080.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OX Burkholderiaceae; Ralstonia.
NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.-C., Cattolico L.,
RA Chandler M., Cholsne N., Claudel-Renaud C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Signier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646065; CAD15388.1; -.
DR InterPro; IPR006453; Gpi6_SPP1.
DR TIGRFAMs; TIGR01563; gpi6_SPP1; 1.
KW Complete proteome.
SQ SEQUENCE 107 AA; 12024 MW; 96CB24F6501459BB CRC64;

Query Match 100.0%; Score 25; DB 6; Length 107;
Best Local Similarity 50.0%; Pred. No. 1e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIYXXDX 8
   :|::||:
Db      34 VEFIELDI 41

RESULT 72
Q8XYS2 PRELIMINARY; PRT; 108 AA.
AC Q8XYS2;
ID Q863N6;
IT Q863N6;
DT 01-MAR-2003 (TrEMBLrel. 20, Created)
DE Probable phage HK022 GP9-related protein.
DN RSC1686 OR RS04080.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OX Burkholderiaceae; Ralstonia.
NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.-C., Cattolico L.,
RA Chandler M., Cholsne N., Claudel-Renaud C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Signier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646065; CAD15388.1; -.
DR InterPro; IPR006453; Gpi6_SPP1.
DR TIGRFAMs; TIGR01563; gpi6_SPP1; 1.
KW Complete proteome.
SQ SEQUENCE 108 AA; 12298 MW; 2D5E38CE1C7E24EC CRC64;

Query Match 100.0%; Score 25; DB 16; Length 108;
Best Local Similarity 50.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIYXXDX 8
   :|::||:
Db      45 KEFIADNR 52

RESULT 73
Q7VIX5 PRELIMINARY; PRT; 108 AA.
ID Q7VIX5;
AC Q7VIX5;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DE Hypothetical protein.
DN RSC1686 OR RS04080.
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OX Helicobacteraceae; Helicobacter.
NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954;
RA Suerbaum S., Josephans C., Sterzenbach T., Drescher B., Brandt P.,
RA Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA Schauer D.B., Shen Z., Weber J.J., Frosch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
Helicobacter hepaticus.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR EMBL; AE017145; AAP7074.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 108 AA; 12836 MW; 93DBA05000CDB51 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 108;
Best Local Similarity 50.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIYXXDX 8
   :|::||:
Db      20 DEFFGYDR 27

```



QY 1 XEFIXDX 8  
:||||:|:  
Db 13 KEFIYDC 20

## RESULT 74

Q83UP9 Q83UP9 PRELIMINARY; PRT; 109 AA.  
AC Q83UP9;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Dinitrogenase reductase Pe protein (fragment).  
GN NIFH.  
OS Gamma-proteobacterium Hot 75m4.  
OC Bacteria; environmental samples.  
OX NCBI\_TaxID=77133;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zehr J.P., Jenkins B.D., Short S.M., Steward G.F.;  
RT "Nitrogenase gene diversity and microbial community structure: a  
RT cross-system comparison."  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY223946; AAC67605.1; -  
DR EMBL; AY223952; AAC67611.1; -  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR000392; Nitrogenase.  
DR Pfam; PF00142; fer4\_NiFe; 1  
DR PRINTS; PR00091; NITROGNASEII.  
DR PROSITE; PS00746; NIFH\_FRXC 1; 1.  
DR PROSITE; PS00692; NIFH\_FRXC 2; 1.  
FT NON TER 1  
FT NON TER 109  
SQ SEQUENCE 109 AA; 11768 MW; 2A3FADB64DBF4269 CRC64;

Query Match 100.0%; Score 25; DB 2; Length 109;  
Best Local Similarity 50.0%; Pred. No. 1.1e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|:  
Db 76 TEFIYDV 83

## RESULT 75

Q8H8H2 Q8H8H2 PRELIMINARY; PRT; 109 AA.  
AC Q8H8H2;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE Hypothetical protein.  
GN QJ1126B12.9  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wing R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Saski C.,  
RA Currie J., Collura K.;  
RT "Rice Genomic Sequence."  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC098695; AN74834.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 109 AA; 12338 MW; C281AC7278D73F29 CRC64;

Query Match 100.0%; Score 25; DB 10; Length 109;  
Best Local Similarity 50.0%; Pred. No. 1.1e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|:  
Db 59 IEFIDKDL 66

## RESULT 76

Q41901 Q41901 PRELIMINARY; PRT; 110 AA.  
AC Q41901;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE NADH ubiquinone oxidoreductase subunit.  
GN Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia; TISSUE=Green siliques;  
RA Raynal M., Grellet F., Laudie M., Meyer Y., Cooke R., Delseny M.;  
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z17587; CAA79002.1; -  
KW Ubiquinone.  
SQ SEQUENCE 110 AA; 12204 MW; A84EDCAC52B9DAF8 CRC64;

Query Match 100.0%; Score 25; DB 10; Length 110;  
Best Local Similarity 50.0%; Pred. No. 1.1e+03;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8  
:||||:|:  
Db 84 IEFICLDL 91

## RESULT 77

Q9M9M6 Q9M9M6 PRELIMINARY; PRT; 110 AA.  
AC Q9M9M6;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE T17B22.24 protein (Hypothetical protein) (At3g03070).  
GN T17B22.24.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,  
RA Renning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
RT "Arabidopsis thaliana chromosome III BAC T17B22 genomic sequence."  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Chao Q., Brooks S., Chen H., Johnson-Hopson C., Khan S., Shinn P.,  
RA Ecker J.;  
RT "Full length cDNA sequence of Arabidopsis thaliana."  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Haas B.J., Volkovsky N., Town C.D., Troukhan M., Alexandrov N.,  
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
RT "Full-length messenger RNA sequences greatly improve genome  
RT annotation."  
RL Genome Biol. 0:0-0(2002).

[4]

```

RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.,
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Hayaishizaki Y., Hsuan V.W.,
RA Ichida J., Jones T., Kamiya A., Karlus-Numann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,
RA Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC012328; AAF26117.1; -
DR EMBL; AF325039; AAG40391.1; -
DR EMBL; AY087471; AAG65015.1; -
DR EMBL; BT004574; AAO42820.1; -
KW Hypothetical protein.
SQ SEQUENCE 110 AA; 12234 MW; 8B81416C52B9C437 CRC64;

Query Match 100.0%; Score 25; DB 10; Length 110;
Best Local Similarity 50.0%; Pred. No. 1.le+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 84 IEFICLDL 91

RESULT 78
Q8G2V8 PRELIMINARY; PRT; 110 AA.
AC Q8G2V8;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Conserved hypothetical protein.
GN BR0204.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Landler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
animal and plant pathogens and symbionts.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR EMBL; AR014333; AAN29155.1; -
DR TTGR; BR0204;
DR InterPro; IPR008497; DUF779.
DR Pfam; PF05610; DUF779; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 110 AA; 11720 MW; 0862D9576C16EF33 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 110;
Best Local Similarity 50.0%; Pred. No. 1.le+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 44 GEFIVGDT 51

us-09-660-302e-1.rspt

RESULT 79
Q9AZU2 PRELIMINARY; PRT; 114 AA.
AC Q9AZU2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Orf30.
GN ORF30.
OS Bacteriophage bL1286.
Virus; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
NCBI_TaxID=151536;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=21113149; PubMed=11160885;
RA Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;
RT "Analysis of six prophages in Lactococcus lactis IL1403: different
genetic structure of temperate and virulent phage populations.";
RL Nucleic Acids Res. 29:644-651(2001).
[2]
RN SEQUENCE FROM N.A.
RA Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., Chopin M.-C.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323669; AAK08317.1; -
SQ SEQUENCE 114 AA; 12457 MW; 137B66755D6DF7AE CRC64;

Query Match 100.0%; Score 25; DB 9; Length 114;
Best Local Similarity 50.0%; Pred. No. 1.le+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 22 GEFIVCDE 29

RESULT 80
Q9CFQ8 PRELIMINARY; PRT; 114 AA.
AC Q9CFQ8;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Prophage p13 protein 34.
GN PI334 OR L1407.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jallou O., Malarne K.,
RA Weissbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL; AF006372; AAK05505.1; -
DR PIR; G86800; G86800.
KW Complete proteome.
SQ SEQUENCE 114 AA; 12458 MW; C5AD66755D6DFF0C CRC64;

Query Match 100.0%; Score 25; DB 16; Length 114;
Best Local Similarity 50.0%; Pred. No. 1.le+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXDX 8
DB 22 GEFIVCDE 29

RESULT 81
Q9LTR0

```

```

ID Q9LTR0 PRELIMINARY; PRT; 115 AA.
AC Q9LTR0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Emb|CAA17159.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Nakamura Y.;
RX MEDLINE=20277480; PubMed=10819329;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:131-135 (2000).
DR EMBL; AB024036; BAB02814.1; -.
SQ SEQUENCE 115 AA; 13512 MW; 81A9D45FFF0B9DC4 CRC64;

Query Match 100.0%; Score 25; DB 10; Length 115;
Best Local Similarity 50.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 19 DEFITEDE 26

RESULT 82
Q8XC01
ID Q8XC01 PRELIMINARY; PRT; 116 AA.
AC Q8XC01;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein z1420.
GN Z1420 OR ECS1157.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Fan C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Fida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22 (2001).

```

```

DR EMBL; AE005294; AAG55549.1; -.
DR EMBL; AP002554; BAB34580.1; ALT_INIT.
DR FIR; A85636; A85636.
DR FIR; E90773; E90773.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 116 AA; 13270 MW; 4B017DE2635A5A6F CRC64;

Query Match 100.0%; Score 25; DB 16; Length 116;
Best Local Similarity 50.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 11 PEFIFNDN 18

RESULT 83
Q7VBN6 PRELIMINARY; PRT; 116 AA.
ID Q7VBN6;
AC Q7VBN6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Uncharacterized protein.
GN PRO1056.
OS Prochlorococcus marinus.
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=1219;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SARG / CCMP 1375 / SS120;
RX MEDLINE=22810154; PubMed=12917486;
RA Dufresne A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,
RA Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,
RA Makarova K.S., Ostrowski M., Oztas S., Robert C., Rogozin I.B.,
RA Scanlan D.J., Tandeau de Marsac N., Weissenbach J., Wincker P.,
RA Wolf Y.I., Hess W.R.;
RT "Genome sequence of the cyanobacterium Prochlorococcus marinus SS120,
RT a nearly minimal ophototrophic genome."
RL Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025 (2003).
DR EMBL; AE017164; AAQ00101.1; -.
KW Complete proteome.
SQ SEQUENCE 116 AA; 13144 MW; 150B1F4124122111 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 116;
Best Local Similarity 50.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 4 PEFIATDN 11

RESULT 84
Q81VI3 PRELIMINARY; PRT; 118 AA.
ID Q81VI3;
AC Q81VI3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN BA0219.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,

```

```

RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.B., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RA "The genome sequence of Bacillus anthracis Ames and comparison to
RT "the genome sequence of Bacillus anthracis Ames and comparison to
RT "closely related bacteria.";
RL Nature 423:81-86(2003).
DR TIGR; BA0219; AAP24263.1; -.
DR InterPro; IPR006542; Cons hypoth1655.
DR TIGRFAMS; TIGR01655; yxeA_fam; 1.
DR Hypothetical protein; Complete proteome.
KW HYPOTHETICAL PROTEIN; Complete proteome.
SQ SEQUENCE 118 AA; 13876 MW; BAE8EA7328AE91CE CRC64;

Query Match 100.0%; Score 25; DB 16; Length 118;
Best Local Similarity 50.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXXDX 8
Db 58 YEFIAFDE 65

RESULT 85
Q81IW3 PRELIMINARY; PRT; 118 AA.
AC Q81IW3;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN BC0233.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelson B.,
RA Kapural V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Gotsman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis.";
RL Nature 423:87-91(2003).
DR EMBL; AB016998; AAP07302.1; -.
DR InterPro; IPR006542; Cons hypoth1655.
DR TIGRFAMS; TIGR01655; yxeA_fam; 1.
DR Hypothetical protein; Complete proteome.
KW HYPOTHETICAL PROTEIN; Complete proteome.
SQ SEQUENCE 118 AA; 13830 MW; C307191648AFB9FA CRC64;

Query Match 100.0%; Score 25; DB 16; Length 118;
Best Local Similarity 50.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXXDX 8
Db 58 YEFIAFDE 65

RESULT 86
Q9KMD8 PRELIMINARY; PRT; 119 AA.
AC Q9KMD8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein VCA0423.
GN VCA0423.

Vibrio cholerae.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004376; AAF96329.1; -.
DR PIR; F82462; F82462.
DR TIGR; VCA0423; -.
DR InterPro; IPR007712; Plasmid stabil.
DR Pfam; PF05016; Plasmid stabil; 1.
DR Hypothetical protein; Complete proteome.
KW HYPOTHETICAL PROTEIN; Complete proteome.
SQ SEQUENCE 119 AA; 13609 MW; 8CC94AB091E3C8C0 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 119;
Best Local Similarity 50.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXXDX 8
Db 17 AEFIALDN 24

RESULT 87
Q8GGN0 PRELIMINARY; PRT; 120 AA.
AC Q8GGN0;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
GN LNMV.
OS Streptomyces atroolivaceus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=66869;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22336326; PubMed=12446651;
RA Cheng Y.Q., Tang G.L., Shen B.;
RT Identification and Localization of the Gene Cluster Encoding
RT Biosynthesis of the Antitumor Macrolactam Leinamycin in Streptomyces
RT atroolivaceus S-140.";
RL J. Bacteriol. 184:7013-7024(2002).
DR EMBL; AF484556; AAN85535.1; -.
KW Hypothetical protein.
SQ SEQUENCE 120 AA; 14105 MW; 95BC6B553E2940D1 CRC64;

Query Match 100.0%; Score 25; DB 2; Length 120;
Best Local Similarity 50.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 XEFIXXDX 8
Db 69 NEFIQTDD 76

RESULT 88
Q8EX15 PRELIMINARY; PRT; 120 AA.
ID Q8EX15
AC Q8EX15;
```



```
DR TIGR: S01413; -.
DR GO: 0005489; P: electron transporter activity; IEA.
DR GO: 0006118; P: electron transport; IEA.
DR InterPro: IPR000345; CytC_heme_BS.
DR PROSITE: PS00190; CYTOCHROME_C; 4.
KW Complete proteome.
SQ SEQUENCE 123 AA; 13307 MW; 560FEIAC95A0C804 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 123;
Best Local Similarity 50.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 37 HEFIYQDG 44

RESULT 92
Q93FUS PRELIMINARY; PRT; 124 AA.
AC Q93FUS;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical dimethyl adenosine transferase (Fragment).
OS Cowdria ruminantium
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Highway;
RX MEDLINE=21472264; PubMed=11587856;
RA Barbet A.F., Whitmire W.M., Kamper S.M., Simbi B.H., Ganta R.R.,
RA Moreland A.L., Mwangi D.M., McGuire T.C., Mahan S.M.;
RT "A subset of Cowdria ruminantium genes important for immune
RT recognition and protection.";
RL Gene 275:287-298 (2001).
DR EMBL; AF388662; AAL08808.1; -.
DR GO: 0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO: 0016740; F: transferase activity; IEA.
DR InterPro: IPR000051; SAM_bind.
KW Transferase.
FT NON TER 124 124
SQ SEQUENCE 124 AA; 14090 MW; 57388DC0CE4975A7 CRC64;

Query Match 100.0%; Score 25; DB 2; Length 124;
Best Local Similarity 50.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 85 YEFILSDA 92

RESULT 93
Q8L3Q2 PRELIMINARY; PRT; 124 AA.
ID Q8L3Q2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OSUNAA0036D19.9 OR OSJNBA0050E08.2.
OS Oryza sativa (Rice), and
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530, 39947;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=O.sativa, and O.sativa (japonica cultivar-group);

Wing R.A., Yu Y., Yang T.J., Nah G., Soderlund C., Chen M., Kim H.-R.,
Rambo T., Saeki C., Henry D., Oates R., Simmons J.;
"Rice Genomic Sequence.";
Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP SPECIES=O.sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;
RC The Rice Chromosome 10 Sequencing Consortium;
"In-depth view of structure, activity, and evolution of rice
chromosome 10.";
RL Science 300:1566-1569 (2003).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=O.sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;
BUELL C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116600; AAN04141.1; -.
DR EMBL; AC098696; AAN04201.1; -.
DR EMBL; AE017059; AAP52225.1; -.
DR Gramene; Q8L3Q2; -.
KW Hypothetical protein.
SQ SEQUENCE 124 AA; 12871 MW; FB3B4ED959E4C1E8 CRC64;

Query Match 100.0%; Score 25; DB 10; Length 124;
Best Local Similarity 50.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 67 AEFISPDH 74

RESULT 94
Q9CLP6 PRELIMINARY; PRT; 124 AA.
ID Q9CLP6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein FM1171.
GN FM1171.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
DR EMBL; AE006157; AAK03255.1; -.
DR InterPro: IPR005220; Cons_hypoth156.
DR Pfam; PF04076; DUF388; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 124 AA; 13402 MW; 289E9736A09BBAA4 CRC64;

Query Match 100.0%; Score 25; DB 16; Length 124;
Best Local Similarity 50.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 69 DEFIFRDV 76

RESULT 95
Q859F7 PRELIMINARY; PRT; 127 AA.
ID Q859F7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
```

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 OS Pseudomonas phage gh-1.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;  
 OC T7-like viruses  
 OX NCBI\_TaxID=197783;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kovalyova I.V., Kropinski A.M.;  
 RT "The complete sequence of Pseudomonas putida bacteriophage gh-1: A  
 RL member of T7 family of viruses.";   
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF493143; AA073159.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 127 AA; 14190 MW; F38DB58355D3A48D CRC64;  
  
 Query Match 100.0%; Score 25; DB 9; Length 127;  
 Best Local Similarity 50.0%; Pred. NO. 1.3e+03;  
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 XEFIXDX 8  
 Db :|||:|:  
 73 REFITKDM 80  
  
 RESULT 96  
 Q89445 PRELIMINARY; PRT; 129 AA.  
 AC Q89445;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE PD129L.  
 GN D129L.  
 OS African swine fever virus (ASFV).  
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.  
 OX NCBI\_TaxID=10497;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=96036500; PubMed=7483270;  
 RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,  
 RA Rodriguez J.F., Vinuela E.;  
 RT "Immune protection conferred by the baculovirus-related glycoprotein  
 RL of Thogoto virus (Orthomyxoviridae).";  
 RT Virology 208:249-278(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=94233765; PubMed=8178480;  
 RA La Vega I., Gonzalez A., Blasco R., Calvo V., Vinuela E.;  
 RT "Nucleotide sequence and variability of the inverted terminal  
 RL repetitions of African swine fever virus DNA.";   
 RL Virology 201:152-156(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=90219205; PubMed=2325203;  
 RA Gonzalez A., Calvo V., Almazan F., Almendral J.M., Ramirez J.C.,  
 RA La Vega I., Blasco R., Vinuela E.;  
 RT "Multigene families in African swine fever virus: family 360.";   
 RL J. Virol. 64:2073-2081(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=90219204; PubMed=2325202;  
 RA Almendral J.M., Almazan F., Blasco R., Vinuela E.;  
 RT "Multigene families in African swine fever virus: family 110.";   
 RL J. Virol. 64:2064-2072(1990).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=91134988; PubMed=1994575;

RA Camacho A., Vinuela E.;  
 RT "Protein p22 of African swine fever virus: an early structural protein  
 RL that is incorporated into the membrane of infected cells.";   
 RL Virology 181:251-257(1991).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX Almazan F., Murguia J.R., Rodriguez J.M., La Vega I., Vinuela E.;  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=94187118; PubMed=8139051;  
 RA Rodriguez J.M., Yanez R.J., Pan R., Rodriguez J.F., Salas M.L.,  
 RA Vinuela E.;  
 RT "Multigene families in African swine fever virus: family 505.";   
 RL J. Virol. 68:2746-2751(1994).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=93346971; PubMed=8393914;  
 RA Yanez R.J., Rodriguez J.M., Rodriguez J.F., Salas M.L., Vinuela E.;  
 RT "African swine fever virus thymidylate kinase gene: sequence and  
 RL transcriptional mapping.";   
 RL J. Gen. Virol. 74:1633-1638(1993).  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=94065656; PubMed=8245848;  
 RA Alcami A., Angulo A., Vinuela E.;  
 RT "Mapping and sequence of the gene encoding the African swine fever  
 RL virion protein of M(r) 11500.";   
 RL J. Gen. Virol. 74:2317-2324(1993).  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=93277388; PubMed=8503790;  
 RA Munoz M., Freije J.M., Salas M.L., Vinuela E., Lopez-Otin C.;  
 RT "Structure and expression in E. coli of the gene coding for protein  
 RL p10 of African swine fever virus.";   
 RL Arch. Virol. 130:93-107(1993).  
 RN [11]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=90357780; PubMed=2389555;  
 RA Blasco R., Lopez-Otin C., Munoz M., Bockamp E.O., Simon-Mateo C.,  
 RA Vinuela E.;  
 RT "Sequence and evolutionary relationships of African swine fever virus  
 RL thymidine kinase.";   
 RL Virology 178:301-304(1990).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=93281390; PubMed=8506138;  
 RA Yanez R.J., Boursnell M., Nogal M.L., Yuste L., Vinuela E.;  
 RT "African swine fever virus encodes two genes which share significant  
 RL homology with the two largest subunits of DNA-dependent RNA  
 RL polymerases.";   
 RL Nucleic Acids Res. 21:2423-2427(1993).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=93353606; PubMed=8102411;  
 RA Rodriguez J.M., Yanez R.J., Almazan F., Vinuela E., Rodriguez J.F.;  
 RT "African swine fever virus encodes a CD2 homolog responsible for the  
 RL adhesion of erythrocytes to infected cells.";   
 RL J. Virol. 67:5312-5320(1993).  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA71V;  
 RX MEDLINE=94085774; PubMed=8262374;  
 RA Yanez R.J., Rodriguez J.M., Boursnell M., Rodriguez J.F., Vinuela E.;  
 RT "Two putative African swine fever virus helicases similar to yeast





```

RX MEDLINE=99376745; PubMed=10447867;
RA Sommer S., Tichy H.;
RT "Major histocompatibility complex (MHC) class II polymorphism and
RT paternity in the monogamous Hypogeomys antimena, the endangered,
RT largest endemic Malagasy rodent.";
RL Mol. Ecol. 8:1253-1272(1999).
DR EMBL; AUI33487; CAB38637.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0045012; F:MHC class II receptor activity; IEA.
DR GO; GO:0019884; P:antigen presentation, exogenous antigen; IEA.
DR GO; GO:0019886; P:antigen processing, exogenous antigen via M. . .; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR001003; MHC_II_alpha.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00993; MHC_II_alpha; 1.
DR SMART; SM00407; IG_C1; 1.
DR PROSITE; PS0835; IG_LIKE; 1.
KW Glycoprotein; MHC II; Transmembrane.
FT NON TER 1
FT NON TER 130
SQ SEQUENCE 130 AA; 14893 MW; 51FA35110E07BECA CRC64;

Query Match 100.0%; Score 25; DB 7; Length 130;
Best Local Similarity 50.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 20 GEFINLDF 27

RESULT 99
Q8GUR8
ID Q8GUR8 PRELIMINARY; PRT; 130 AA.
AC Q8GUR8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Thioredoxin h. (Garden pea).
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
CX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RA Montrichard F., Renard M., Duval F.D., Macherel D.;
RT "Expression of the NADP-thioredoxin reductase/thioredoxins h system
RT during germination of seeds of Pisum sativum L. ";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV170651; AA012855.1; -.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR006662; Thiored.
DR InterPro; IPR005746; Thioredoxin.
DR InterPro; IPR006663; Thioredox_dom2.
DR Pfam; PF00085; thiored; 1.
DR PRINTS; PR00421; THIOREDOXIN.
DR TIGRFAMs; TIGR01068; thioredoxin; 1.
DR PROSITE; PS00194; THIOREDOXIN; 1.
SQ SEQUENCE 130 AA; 14550 MW; 4088F65F1BECF5AE CRC64;

Query Match 100.0%; Score 25; DB 10; Length 130;
Best Local Similarity 50.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 76 VEFIKIDV 83

```

```

RESULT 100
Q39MY7
ID Q39MY7 PRELIMINARY; PRT; 130 AA.
AC Q39MY7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Growth hormone receptor (Fragment).
OS Meriones unguiculatus (Mongolian jird) (Mongolian gerbil).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC Meriones.
CX NCBI_TaxID=10047;
RN [1]
RP SEQUENCE FROM N.A.
RA Fathimani K., Hull K., Harvey S.;
RT "Sequence analysis of the intracellular domain of rodent growth
RT hormone receptor cDNAs ";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF247184; AAK28386.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON TER 1
FT NON TER 130
SQ SEQUENCE 130 AA; 14589 MW; 28CD401E9E876F2D CRC64;

Query Match 100.0%; Score 25; DB 11; Length 130;
Best Local Similarity 50.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XEFIXDX 8
Db 35 VEFIELDI 42

```

Search completed: September 12, 2004, 02:38:06  
Job time : 95 secs

**This Page Blank (uspto)**